TEXTE
41/2024

Critical Evaluation of Effect Models for the Risk Assessment of Plant Protection Products

Final Report

For our Environment

Umwelt Bundesamt

TEXTE 41/2024

Environmental Research of the
Federal Ministry for the
Environment, Nature Conservation
and Nuclear Safety

Project No. (FKZ) 3715 67 408 0
FB001318/ENG

Critical Evaluation of Effect Models for the Risk Assessment of Plant Protection Products

by

Jeremias Becker, Matthias Liess
Helmholtz Centre for Environmental Research GmbH – UFZ, Leipzig

Stephanie Kramer-Schadt, Mathias Franz
Leibniz-Institut für Zoo- und Wildtierforschung, Berlin

Tjalling Jager
DEBtox Research, Stevensweert

On behalf of the German Environment Agency

Imprint

Publisher:
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
Internet: www.umweltbundesamt.de

Study led by:
Helmholtz-Zentrum für Umweltforschung GmbH – UFZ
Department System-Ökotoxikologie
Permoserstraße 15
04318 Leipzig

Study completed in:
August 2022

Edited by:
Section IV 1.3 Plant Protection Products
Magali Solé, Sabine Duquesne, Silvia Pieper*

*present: II 2.7 Soil State and Soil Monitoring

Publication as pdf:
http://www.umweltbundesamt.de/publikationen

ISSN 1862-4804

Dessau-Roßlau, March 2024

Abstract

Mechanistic effect models have become increasingly popular for use in the frame of the environmental risk assessment of plant protection products and the active substances therein (summarized as pesticides). In 2018, the EFSA Panel on Plant Protection Products and their Residues (PPR) considered TKTD models of the GUTS family. fit for the modelling of acute mortality from pesticide exposure in aquatic ecosystems. Additionally, specialized dynamic energy budget models (DEBtox) for sublethal effects and a number of population models are currently under development for use in Higher Tier studies for risk assessment. The use of such models is expected to increase in the near future, but leaves numerous open questions to risk assessors, modelers, applicants and the public. In this report we thus provide a scientific evaluation of mechanistic effect models and their use, especially in the context of their implementation in the environmental risk assessment (ERA) of pesticides.

In Part 1 of the report, we briefly reviewed a number of ecological models that may be potentially interesting for use in ERA. In part 2, we evaluated in detail 11 models or model families that may be potentially suitable or have been already proposed by applicants for ERA. In part 3, we evaluated some applications of these models that have been submitted for ERA, usually as a refinement tool to demonstrate low risk of unacceptable effects. These case studies were provided by the German Federal Environmental Agency. Part 4 provides a general discussion on conclusions that can be drawn from the evaluations, along with some suggestions for future improvements in model development and their use in ERA. Results have been discussed at an international symposium on 19th and 20th of September 2019 in Berlin with experts in the field of work; the presentations and minutes are presetned in the annex.

Kurzbeschreibung

Mechanistische Modelle für die Vorhersage von Effekten erhalten eine zunehmende Bedeutung in der ökologischen Risikobewertung von Pflanzenschutzmitteln und den darin enthaltenen Wirkstoffen (zusammengefasst als Pestizide bezeichnet). 2018 bescheinigte das EFSA Gremium für Pflanzenschutzmittel und ihre Rückstände (PPR) TKTD –Modellen der GUTS-Familie die Anwendungsreife für die Modellierung akuter Mortalität in aquatischen Ökosystemen infolge einer Pestizidexposition. Darüber hinaus befinden sich spezielle Energiehaushaltsmodelle für die Vorhersage von subletalen Effekten (DEBtox) sowie Populationsmodelle für Higher Tier Studien in der Entwicklung. Die absehbare Zunahme von Modellierungsstudien in naher Zukunft wirft eine Reihe von Fragen für Risikobewerter, Modellierer, Antragsteller und die Öffentlichkeit auf. Der vorliegende Bericht bietet daher eine wissenschaftliche Begutachtung mechanistischer Effektmodelle und ihrer Anwendung, insbesondere im Rahmen der ökologischen Umweltrisikobewertung (ERA) von Pestiziden.

Im ersten Teil des Berichts bieten wir einen kurzen Überblick über eine Reihe von ökologischen Modellen, welche potentiell für die Umweltrisikobewertung interessant erscheinen. Im zweiten Teil wurden 11 Modelle bzw. Modellfamilien detailliert begutachtet, welche potenziell geeignet sind oder bereits von Antragstellern für die Risikobewertung vorgeschlagen wurden. Im dritten Teil wurden einige Anwendungsbeispiele dieser Modelle begutachtet, welche im Rahmen der Zulassungsverfahren eingereicht wurden, üblicherweise zur verfeinerten Risikobewertung, mit der ein geringes Risiko von unakzeptablen Effekten gezeigt werden soll. Die Studien wurden durch das Umweltbundesamt bereitgestellt. Der vierte Teil bietet eine allgemeine Diskussion mit Schlussfolgerungen aus den Evaluationen sowie Vorschläge für die zukünftige Entwicklung und Anwendung von Modellen in der Umweltrisikobewertung. Die Ergebnisse wurden am 19. Und 20. September 2019 auf einem internationalen Symposium in Berlin mit Experten des Fachgebiets diskutiert; die Präsentationen und das Protokoll sind im Anhang verfügbar.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

a.s.	Active substance
AF	Assessment factor
BCF	Bioconcentration factor
DEB	Dynamic energy budget
ECx	Effective concentration (causing x % effect)
EFSA	European Food Safety Authority
ERA	Environmental risk assessment
ERO	Ecological recovery option
ETO	Ecological threshold option
FOCUS	FOrum for the Coordination of pesticide fate models and its USe
GAP	General application patterns
IBM	Individual-based model
IT	Individual tolerance
K_{ow}	*n*-Octanol / water partitioning coefficient
LC50	Lethal median concentration (concentration that kills 50 % of a population)
LoEP	List of endpoints (from EFSA conclusion on a given active substance)
MoA	Mode of action
NOAEL	No observable adverse effect level
ODE	Ordinary differential equation
PDE	Partial differential equation
PEC	Predicted environmental concentration
PPP	Plant protection product
PPR	Panel on Plant Protection Products and their Residues
RAC	Regulatory acceptable concentration
RMS	Rapporteur member state
SD	Stochastic death
SPG	Specific protection goal
TD	Toxicodynamic
TER	Toxicity exposure ratio
TK	Toxicokinetic
TU	Toxic unit (ratio of pesticide concentration vs. LC50)
TWA	Time weighted average
zRMS	Zonal rapporteur member state

Summary

Background

Commercial agriculture in the European Union constantly relies on the intensive use of chemical plant protection products (Eurostat 2020). However, plant protection products (PPPs) and the active substances (a.s.) therein, summarized as pesticides from here on, must be registered in the member states of the European Union (Regulation (EC) No 1107/2009). The registration procedure requires an environmental risk assessment (ERA) in which it must be demonstrated that no unacceptable effects on non-target species may occur from the proposed pesticide use (Commission Regulation (EU) No 546/2011). The risk assessment follows a tiered approach that starts with Tier 1 based on mandatory tests according to the data requirements (Commission Regulation (EU) No 283/2013, No 284/2013). This initial step is considered highly conservative but very general, and can be followed by various options for subsequent refined assessment (Higher Tier studies) if a pesticide did not pass Tier 1 (Commission Regulation (EU) No 546/2011).

The environmental risk assessment (ERA) considers both the environmental fate of a pesticide and the effects it may cause in exposed organisms. Mechanistic simulation models for the environmental fate have been already established in the ERA of pesticides for many years (Richter et al. 1996). However, in the last decade, also an increasing number of ecotoxicological effect models have been developed and applied for the risk assessment of pesticides. These models mechanistically simulate effects of pesticides at various levels of biological organization from the individual to the population and even the ecosystem; they are mostly intended to complement or even replace the established Higher Tier studies, i.e. to be used as a refinement option in cases when unacceptable risk is identified at lower tiers under conservative assumptions. Effect modelling may potentially circumvent limitations of the established experimental approaches. This includes the unrealistic exposure profiles in Tier 1 ecotoxicological tests, and constraints on the number of environmental scenarios and on the duration of effects that may be tested in Higher Tier studies.

Some effect model applications have been already submitted by applicants in the process of regulatory ERA. However, the increasing interest of applicants in ecotoxicological effect modelling provides new challenges for risk assessors and modellers. Key issues involve the realism and the level of protection that can be achieved when using effect models. In 2014, the EFSA Panel on Plant Protection Products and their Residues (PPR) published a Scientific Opinion on Good Modelling Practice (Sci. Op. on GMP). This document provides some guidance on the development, testing and application of mechanistic effect models for the ERA of PPPs (EFSA PPR 2014b). The issues are discussed in a general way with a focus on population (and potentially community and ecosystem) models, but provided general recommendations rather than specific criteria for model development, application and evaluation. This was followed by a Scientific Opinion with a specific review on TKTD models for organism- (= individual-) level effects in aquatic species (EFSA PPR 2018).

In the present report, we provide a scientific evaluation on the state of effect models for the risk assessment of pesticides. In part 1, we reviewed various ecological models that have been published in the scientific literature until 2017 and may be potentially interesting for use in the ERA of pesticides. Due to the high number of models available, this selection cannot cover all potentially relevant models existing.

In part 2, we evaluated 11 selected models in detail based on the available scientific literature and model documentation from 2007 to 2020. The selection covers the GUTS and the DEB framework for lethal and sublethal individual-level effects, respectively. Additionally, we evaluated three population models for freshwater invertebrates (IDamP, MASTEP, and the IBM *Chaoborus* Population Model), one population model for soil-dwelling springtails (SpringSim), and three population models for small mammals (an application of the ALMaSS framework, eVole, and a model for the wood mouse from Liu

et al.). Finally, we evaluated a potential application of the $SPEAR_{pesticides}$ approach for the prediction of effects on a whole freshwater macroinvertebrate community, and the aquatic ecosystem model AQUATOX. Each evaluation starts with a general introduction and overview on the status of a model in terms of development and application. This is followed by a detailed model description and an assessment based on checklists provided in the EFSA Sci. Op. on GMP(2014b) for the documentation by modellers and for the evaluation by risk assessors. Finally, each evaluation is followed by a qualitative assessment of uncertainties that may result in a potential under- or overestimation of the real risk of pesticides.

In part 3, we evaluated a number of case studies from 2007 to 2017 in which some of the models named above were applied to specific pesticides and proposed for the regulatory risk assessment. Here we used again the checklist for model assessment provided in the EFSA Sci. Op. on GMP (2014b).

Finally, in part 4 we provide a discussion on general conclusions that can be drawn on the status of mechanistic effect models for the risk assessment of pesticides. Some results of this report have been presented and discussed on a two-day workshop held in Berlin on 19th and 20th of September 2019 with scientists from academia and contract research, risk assessors from various EU member states, and members of the chemical industry. Presentations and minutes of the workshop are provided in annex 1 and 2.

In general, we suggest to separate the evaluation of a model in general from the evaluation of a specific model application. Mechanistic effect models are typically not developed only for a single application, and higher-level models for populations, communities and ecosystems typically come with a default built-in parameterization for the basic biological part of a model. The general model design and the basic parameterization may be assessed once for the model in general and not re-assessed again for each specific model application, in contrast to the case-specific model setup to a given environmental scenario and the modelling outcome. However, the checklists from the EFSA Sci. Op. on GMP (2014b) tend to mix information on a model in general and on a specific model application. We suggest to revise them accordingly.

Individual-level models

For TKTD models on individual-level effects we identified basically three potential forms of application in the framework of ERA. First, the models may be simply used as alternatives to the traditional static models for the fitting of dose-response curves. The advantage of GUTS or DEB models is that they can use all available data (i. e., from repeated observations after different exposure times on the same replicate, if available); they may thus estimate simple summary statistics like LCx and ECx with higher precision than classical dose-response models that may handle only data from a single observation time point. Additionally, these models provide a mechanistic theoretical background for the interpolation to effects from different test concentrations and at different observation times that is not given for interpolations from classical dose-response models.

Second, TKTD models may be used to extrapolate effects to more realistic exposure profiles that differ from those in standard tests used for model calibration, as described in the Tier 2C approach for refined exposure in the Aquatic Guidance Document (EFSA PPR 2013). However, beforehand, the ability of a model to extrapolate effects must be validated with independent data on a different exposure profile not used for model calibration (EFSA PPR 2018). The evaluated case studies reported such validations, although the provided information was sometimes very limited.

Third, TKTD models may serve as building blocks for the input effects that are imposed at organism level in higher-level models for populations, communities and ecosystems. We call these building blocks toxicity modules. Principally, the same validation criteria as for the Tier 2C approach should apply to this use of TKTD models, since also the exposure profiles that organisms experience in higher-level models typically deviate from those used for calibration of the toxicity module. However, we

identified only a single case study where a TKTD module for a higher-level model has been validated; this case study came from a scientific model demonstration and not from a modelling study actually proposed for ERA.

Population models

Population models can be used in the European framework of ERA to address basically two specific protection goals SPG): First, these models can be applied to assess the potential of a population to recover from acute (usually lethal) effects that considerably decrease its abundance or biomass within a short time (hours to days). Second, population models can be applied to assess whether chronic effects at organism level may result in long-term repercussions on the abundance or biomass of a population. Chronic effects are often sublethal and will not instantly affect population size but act on its individuals over an extended period of time from days to weeks; they may include both after-effects from short-term (acute) exposure and effects from an ongoing chronic exposure.

Accordingly, the potentially most relevant modelling output for risk assessment is the predicted population recovery time and the exposure level at which no long-term decrease in abundance / biomass can be observed. However, a clear definition on when a population has been actually recovered or what extent of population decline is unacceptable has not been established.

Population modelling studies in the ERA of pesticides have been generally justified with the possibility of increased realism due to the inclusion of population recovery through reproduction and recolonization. Without these processes, the real risk of pesticides may be indeed overestimated. However, a risk assessment refined this way may increase realism only when also additional processes are considered that likely decrease the potential of population recovery. Unfortunately, the high level of detail spent on population recovery processes in the models is contrasted by the little attention payed to such limiting processes.

E. g., in most cases the extent of input effects that are imposed on organism-level are likely not representing the true risk: The evaluated modelling studies in part 3 of this report that addressed population recovery from an acute population decline simulated only lethal but no sublethal effects. However, exposure to pesticides that causes acute mortality is likely associated with sublethal and often chronic effects in surviving individuals, such as a decrease in growth, reproduction, competitive strength, and in the ability to escape from predation. Such a decrease in fitness may considerably delay population recovery (Desneux et al. 2006). Additionally, stressors experienced in the field but not in a standard test environment, such as food limitation or temperature stress, may increase the sensitivity of individuals to the simulated input effects of pesticides at organism level in the field (Liess et al. 2016a). Toxicity modules that may consider such interactions of organism-level effects with the environment are rare, and those that are in principal capable of doing so (e.g., DEB) require further validation studies on this particular aspect.

Additionally, environmental stressors such as unfavourable temperature and dissolved oxygen levels in water or predatory and competing species may decrease population growth and thus the potential for population recovery (Liess et al. 2013). It has been shown that the predicted recovery in population models may substantially change with the inclusion of antagonistic species (Gabsi et al. 2014d, Kattwinkel and Liess 2014). However, antagonistic species have never been explicitly simulated and abiotic stressors have been considered only to a minimum extent in the model applications reviewed in part 3 of this report. This is of particular concern when biological processes have been parameterized using laboratory data under comparably favourable conditions. In contrast, where the biological parameterization was based on (semi-)field data, environmental stressors may have been implicitly included in the model to a certain extent. However, even then, effects may have been underestimated because the potentially lowered fitness due to chronic effects of pesticides (see above) may increase the susceptibility of individuals to additional environmental stressors (Becker and Liess 2015).

Taken together, we identified a tendency towards the consideration of those ecological processes that may decrease the risk of pesticides, resulting in a high likelihood of biased outcome from population models that can lead to an underestimation of the real risk.

Community and ecosystem models

General protection goals for communities and ecosystems refer to no unacceptable effects on biodiversity and ecosystem functioning, but no specific protection goals have been established (Commission Regulation (EU) No 546/2011). The lack in SPG for communities and ecosystems, together with the high complexity in parameterization and validation, are major obstacles for higher-level effect modelling in the ERA of pesticides. Accordingly, we found no applications of community and ecosystem models in the prospective regulatory risk assessment so far.

The ecosystem model AQUATOX has been explicitly designed to consider effects of additional stressors on population recovery, as well as indirect effects via the food web and the potential for biomagnification. The toxicity module for input effects at population level can consider acute and chronic effects but no interactions with additional stressors (multiple stressors act in an additive way at organism level). The toxicity module has been designed in a way that requires only a minimum of ecotoxicological information from standard Tier 1 tests. Otherwise, it might not be possible to parameterize a full ecosystem model with numerous species. However, as a result, the toxicity module depends on a number of assumptions like Habers' rule that will not always hold and present an important source of uncertainty in the model. This is probably the main reason why AQUATOX and other ecosystem models are generally used to retrospectively analyse observed effects in the field, but not for prospective risk assessment.

The $SPEAR_{pesticides}$ approach has been originally developed as a bioindicator for the assessment of pesticide pollution in small streams, based on characteristic changes in the freshwater macroinvertebrate community (Liess and von der Ohe 2005). $SPEAR_{pesticides}$ is thus a tool for monitoring but also predicting pesticide effects in the field. The $SPEAR_{pesticides}$ approach may be applied retrospectively to test the actual protectiveness of the established methods in ERA. Additionally, the approach can be used prospectively to predict community-level effects in the field (in terms of SPEAR values). This is done by extrapolating individual-level effects in the laboratory (in terms of toxic units, TU) to community-level effects in the field (in terms of SPEAR values), based on an empirically established TU vs. SPEAR relationship. SPEAR values relate to the ratio of individuals belonging to vulnerable vs. non-vulnerable taxa. They are thus not directly applicable to address the specific SPGs of EFSA for individuals or single populations but highly relevant to immediately address the actual protection goal of high certainty against long-term repercussions on the abundance and diversity of aquatic invertebrates (Regulation (EC) No 1107/2009 , EFSA PPR 2013). The $SPEAR_{pesticides}$ approach may not be customized to the assessment of a specific pesticide with a specific application pattern and mode of action. Nevertheless, $SPEAR_{pesticides}$ relates effects to the maximum TU, and the TU can be predicted for a specific pesticide from fate modelling. Therefore, $SPEAR_{pesticides}$ may be used to screen for effects that are typically expected from a pesticide with a given toxicity under realistic conditions in the field; the predicted effects may serve as a benchmark for potentially exonerating case-specific studies.

Potential for improvement

The development and application of effect models for the risk assessment of pesticides was generally well documented and followed the ODD protocol (overview, design concepts, detail; Grimm et al. 2006, Grimm et al. 2010) or the TRACE framework (TRAnsparent and Comprehensive Ecological model documentation; Schmolke et al. 2010, Grimm et al. 2014) framework that has been proposed for model documentation. However, the EFSA Sci. Op. on GMP (2014b) attaches great value to the evaluation of a

model by means of sensitivity and uncertainty analysis and of validation before a model may be applied for regulatory risk assessment. For TKTD models, a sensitivity analysis may be only of limited use because these models have been fully calibrated to specific data. More relevant information may be obtained when the model fit to the data used for calibration is assessed. In contrast, sensitivity analyses are highly relevant for population models. Unfortunately, the most relevant endpoints from a regulatory point of view, i. e. the sensitivity of predicted recovery times or NOAEL to various input parameters, and particularly to the magnitude of organism-level input effects, have not been addressed in any of the models reviewed in detail.

Additionally, model validation should be considerably improved. Some population models were applied for risk assessment although model predictions had never been matched with independent real-world data at all. In other cases, only model predictions on population dynamics in control scenarios without pesticide exposure have been tested, while the regulatory relevant endpoints (recovery time or NOAEL) were not. Only in one case study, predicted dynamics of an exposed population were compared with "real-world" data (i. e., experimental data from a mesocosm study); however, no long-term effects were observed in this experimental study, so that the potential of the model to identify existing unacceptable effects could not be assessed. The frequently poor extent of model validation may be explained by a general lack of available real-world data to which a population model might be applied. However, the ecological mechanisms in these models, apart from the toxicity module for input effects, are not specific for any type of toxicant. Therefore, we suggest to search for historical data on a variety of toxicants that may be used for model validation. For better comparison / validation, a relation of the increase in population effects (model output) with individual-level effects (model input) may be established and compared with the individual-level vs. population-level effect relationship observed in the real-world data for validation.

In conclusion, mechanistic effect models may help in extrapolating effects from artificial tests to field conditions; they may thus potentially increase realism in the ERA of pesticides. However, care should be taken that the model design is balanced in terms of processes that may decrease or increase the actual risk, and the evaluation of models should be improved. Finally, models can only help to assess the risk of effects that are known *a priori*, but cannot detect potential risk from novel mode of actions. Effect models may therefore complement but not replace experimental work and field monitoring.

Zusammenfassung

Hintergrund

Die kommerzielle Landwirtschaft in der Europäischen Union beruht nach wie vor auf einer intensiven Nutzung von chemischen Pflanzenschutzmitteln (Eurostat 2020). Pflanzenschutzmittel (plant protection products, PPPs) sowie die darin enthaltenen Wirkstoffe (active substances, a.s.), von hier an zusammenfassend als Pestizide bezeichnet, sind in den Mitgliedsstaaten der Europäischen Union zulassungspflichtig (Verordnung (EG) Nr. 1107/2009). Der Zulassungsprozess sieht eine Bewertung des Risikos für die Umwelt (ERA) vor, um sicherzustellen, dass von der empfohlenen Anwendung von Pestiziden keine inakzeptablen Effekte auf Nichtzielorganismen ausgehen werden (Verordnung (EU) Nr. 546/2011 der Kommission). Dabei folgt die Risikobewertung einem gestuften Ansatz, ausgehend von Tier 1, welches auf in den Datenanforderungen (Verordnung (EU) No 283/2013, Verordnung (EU) No 284/2013 der Kommission) vorgeschriebenen Tests basiert. Dieser einleitende Schritt gilt als sehr konservativ aber wenig spezifisch. Erfüllt ein Pestizid die notwendigen Kriterien in diesem Schritt nicht, besteht anschließend die Möglichkeit, verschiedene Verfeinerungen vorzunehmen (Higher Tier-Studien, Verordnug (EG) Nr. 546/2011).

Die Umweltrisikobewertung (environmental risk assessment, ERA) berücksichtigt dabei sowohl den Verbleib eines Pestizids in der Umwelt, als auch die möglichen Effekte in exponierten Organismen. Mechanistische Simulationsmodelle für den Verbleib in der Umwelt haben sich bereits seit vielen Jahren in der Risikobewertung durchgesetzt (Richter et al. 1996). Im vergangenen Jahrzehnt folgte nun auch die Entwicklung und Anwendung einer zunehmenden Zahl an ökotoxikologischen Effektmodellen für die Risikobewertung. Diese mechanistischen Modelle simulieren die Effekte von Pestiziden auf verschiedenen biologischen Organisationsebenen vom Individuum über die Population bis hin zum Ökosystem. Sie sind hauptsächlich zur Unterstützung von oder sogar als Ersatz für Higher Tier-Studien vorgesehen, d. h. als eine Option für eine verfeinerte Risikobewertung, wenn unter konservativen Annahmen in einer Lower Tier-Bewertung ein unannehmbares Risiko festgestellt wurde. Die Modellierung von Effekten bietet eine Möglichkeit, Einschränkungen der etablierten experimentellen Ansätze zu umgehen. Dies betrifft beispielsweise die unrealistischen Expositionsprofile in ökotoxikologischen Tier 1-Tests, sowie die begrenzte Zahl an möglichen Umweltszenarien und die begrenzte Versuchsdauer in Higher Tier-Studien.

Einige Anwendungen von Effektmodellen wurden bereits in Zulassungsverfahren für die Umweltrisikobewertung von Pestiziden eingereicht. Das steigende Interesse der Antragsteller an der Modellierung ökotoxikologischer Effekte bringt jedoch neue Herausforderungen für Risikobewerter und Modellierer mit sich. Strittige Punkte sind u. a. der Realismus und das Schutzniveau, welches mit derartigen Modellen erreicht werden kann. 2014 veröffentlichte das EFSA Gremium für Pflanzenschutzmittel und ihre Rückstände (PPR) ein wissenschaftliches Gutachten (Scientific Opinion) zur guten fachlichen Praxis in der Effektmodellierung für die Bewertung des Umweltrisikos von Pflanzenschutzmitteln (EFSA PPR 2014b). Hier wurden die Entwicklung, das Testen und die Anwendung von Modellen im Allgemeinen ausführlich diskutiert, mit einem besonderen Fokus auf Populationsmodellen (und potentiellen Modellen für Lebensgemeinschaften und Ökosystemen). Das Dokument enthält allerdings wenig greifbare Kriterien für Entwickler, Anwender und Risikobewerter. Später folgte ein spezifisches Gutachten zu TKTD Modellen für Effekte auf Organismus- (= Individuen-) Ebene für aquatische Lebewesen (EFSA PPR 2018).

Im vorliegenden Projektbericht stellen wir ein wissenschaftliches Gutachten zum Stand der Entwicklung und Anwendung von Effektmodellen in der Risikobewertung vor. Der erste Teil gibt einen kurzen Überblick über einige bis zum Jahr 2017 in der Fachliteratur verfügbare ökologische Modelle, welche

sich potentiell für die Bewertung des Umweltrisikos von Pestiziden eignen. Aufgrund der großen Anzahl von Modellen kann diese Auswahl keinen vollständigen Überblick über sämtliche potentiell relevante Modelle geben.

Im zweiten Teil wurden 11 ausgewählte Modelle basierend auf den von 2007 bis 2020 verfügbaren wissenschaftlichen Publikationen und Modell-Dokumentationen im Detail begutachtet. Diese Auswahl umfasst die GUTS- und DEB-Familien für die Modellierung akuter und chronischer Effekte auf Individuen-Ebene. Des Weiteren wurden drei Populationsmodelle für Wirbellose im Süßwasser (IDamP, MASTEP und das IBM *Chaoborus* Population Model), ein Populationsmodell für im Boden lebende Springschwänze (SpringSim) sowie drei Populationsmodelle für Kleinsäuger (eine Anwendung von ALMaSS, eVole und ein Modell für die Waldmaus von Liu et al.) evaluiert. Schließlich wurde eine mögliche Anwendung des $SPEAR_{pesticides}$-Ansatzes für die Vorhersage von Effekten auf die Lebensgemeinschaft von Makroinvertebraten in kleinen Fließgewässern, sowie das aquatische Ökosystemmodell AQUATOX evaluiert. Jedes Gutachten beginnt mit einer allgemeinen Einleitung zum Stand der Entwicklung und Anwendung eines Modells. Darauf folgt zunächst eine detaillierte Beschreibung und anschließend eine Bewertung basierend auf Fragebögen für Modellierer und Risikobewerter aus dem EFSA Gutachten zur guten Modellierungs-Praxis (2014b). Abschließend erfolgt jeweils eine qualitative Bewertung der Unsicherheiten im Zusammenhang mit einem Modell, die zu einer potentiellen Unter- oder Überschätzung des tatsächlichen Risikos von Pestiziden führen können.

Im dritten Teil wurden eine Reihe von Fallstudien von 2007 bis 2017 begutachtet, in denen einige der o. g. Modelle zur Risikobewertung spezifischer Anwendungen von Pestiziden eingereicht wurden. Dazu wurden wieder die Fragen zur Modellbewertung aus dem EFSA Gutachten zur guten Modellierungs-Praxis verwendet (2014b).

Der vierte Teil umfasst schließlich eine Diskussion zu allgemeinen Schlussfolgerungen aus den vorangegangenen Begutachtungen bez. des derzeitigen Stands mechanistischer Effektmodelle für die Risikobewertung. Einige dieser Ergebnisse wurden auf einem zweitägigen Workshop am 19. Und 20. September 2019 in Berlin vorgestellt und mit Wissenschaftlern aus der universitären sowie der Auftragsforschung, Risikobewertern aus verschiedenen EU Mitgliedsstaaten sowie Vertretern der chemischen Industrie diskutiert. Die Präsentationen und das Protokoll sind in den Anhängen 1 und 2 verfügbar.

Im Allgemeinen empfehlen wir eine Trennung zwischen der Begutachtung eines Modells an sich sowie von spezifischen Anwendungen des Modells. Mechanistische Effektmodelle werden üblicherweise nicht für eine einzige Anwendung entwickelt, und höherstufige Modelle für Populationen, Lebensgemeinschaften und Ökosysteme beinhalten i. d. Regel eine standardmäßige Parametrisierung für den physiologischen Teil eines Modells. Es erscheint im Allgemeinen ausreichend, das generelle Design und die eingebaute Standard-Parametrisierung eines Modells einmal zu bewerten. Im Gegensatz dazu ist die Einstellung des jeweiligen Umweltszenarios fallspezifisch und sollte wie die Modellierungsergebnisse für jede Anwendung separat begutachtet werden. Da die Fragebögen aus dem EFSA Gutachten zur guten Modellierungs-Praxis (2014b) tendenziell Informationen zu einem Modell im Allgemeinen und zu einer spezifischen Anwendung vermischen, empfehlen wir eine entsprechende Überarbeitung dieser Fragebögen.

Modelle für Individuen

Für TKTD-Modelle für Effekte auf Individuen-Ebene wurden im Wesentlichen drei potenzielle Anwendungsformen im Rahmen der regulatorischen Bewertung des Umweltrisikos identifiziert. Zum einen können diese Modelle schlicht als Alternativen zur traditionellen statischen Modellierung von Dosis-Wirkungskurven dienen. Der Vorteil von GUTS- oder DEB –Modellen liegt dabei in der Möglichkeit, sämtliche Daten, d. h. auch ggf. vorhandene wiederholte Beobachtungen nach verschiedenen Expositionszeiten am selben Replikat, zu verwenden. Dies ermöglicht eine genauere Schätzung von Dosis-Wir-

kungsbeziehungen im Vergleich zu klassischen Modellen, welche nur Daten eines einzelnen Beobachtungszeitraums verwenden können. Zudem bieten GUTS- und DEB-Modelle einen mechanistischen theoretischen Hintergrund für die Interpolation von Effekten zwischen verschiedenen Konzentrationen und Beobachtungszeiträumen, welches für klassische Dosis-Wirkungsmodelle nicht existiert.

Des Weiteren können TKTD-Modelle für Extrapolationen verwendet werden, um Effekte von Expositionsprofilen vorherzusagen, die realistischer im Vergleich zu den Standardtests sind, mit welchen sie kalibriert wurden. Insbesondere können diese Modelle gemäß dem im Leitfaden für die aquatische Risikobewertung (EFSA PPR 2013) beschriebenen Tier 2C-Ansatz verwendet werden, um die Effekte eines verfeinerten Expositionsprofils zu simulieren. Allerdings muss zunächst die Fähigkeit eines kalibrierten Modells zur Extrapolation validiert werden. Dies geschieht durch den Vergleich der Modellvorhersagen mit einem unabhängigen Datensatz aus einem anderen Expositionsprofil (EFSA PPR 2018). Die Fallstudien enthalten entsprechende Validierungen, allerdings teilweise mit zu wenig geeigneten Informationen.

Schließlich können TKTD-Modelle als Bausteine für die Eingangseffekte auf der Ebene der Organismen in höherstufigen Modellen für Populationen, Lebensgemeinschaften und Ökosystemen dienen. Diese Bausteine werden im weiteren Text als Toxizitätsmodule bezeichnet. In diesem Fall sollten grundsätzlich die gleichen Validitätskriterien wie für den Einsatz im Tier 2C-Ansatz gelten, da auch in höherstufigen Modellen das Expositionsprofil für die einzelnen Organismen von dem Expositionsprofil für die Kalibrierung des Moduls abweicht. Allerdings wurde lediglich eine einzige Studie identifiziert, in welcher ein Toxizitätsmodul validiert wurde, und diese diente wissenschaftlichen Demonstrationszwecken und nicht einer tatsächlichen spezifischen Risikobewertung.

Populationsmodelle

Die Nutzung von Populationsmodellen im Rahmen der europäischen regulatorischen Risikobewertung bezieht sich im Wesentlichen auf zwei spezifische Schutzziele (specific protection goals, SPG): Zum einen kann mit diesen Modellen das Potenzial zur Wiedererholung einer Population nach akuten (i. d. Regel letalen) Effekten eingeschätzt werden, welche in kurzer Zeit (Stunden bis Tage) die Abundanz oder Biomasse einer Population deutlich verringern. Zum anderen kann mit Populationsmodellen geschätzt werden, inwiefern chronische Effekte auf Individuen-Ebene, zu langfristigen Folgen für die Abundanz / Biomasse einer Population führen. Chronische Effekte sind i. d. Regel subletal und beeinträchtigen eine Population nicht unmittelbar, wirken aber über einen ausgedehnten Zeitraum von Tagen bis Wochen; sie umfassen sowohl Nachwirkungen einer kurzfristigen (akuten) Exposition als auch die unmittelbaren oder verzögerten Folgen einer anhaltenden chronischen Exposition.

Die aus regulatorischer Sicht relevantesten Endpunkte der Modellierung sind dementsprechend die vorhergesagte Wiedererholungszeit sowie der Expositionsgrad, bei welchem keine Langzeiteffekte auf die Abundanz beobachtet werden können. Allerdings fehlt bisher eine eindeutige Definition darüber, ab wann eine Population als wiedererholt gilt bzw. welches Ausmaß an Langzeiteffekten inakzeptabel ist.

Studien mit Populationsmodellen zur Bewertung des Umweltrisikos von Pestiziden werden im Allgemeinen mit der Möglichkeit begründet, durch die Berücksichtigung der Wiedererholung von Populationen aufgrund von Reproduktion und Wiederbesiedlung einen höheren Grad an Realismus zu erzielen. Ohne solche Prozesse kann das Risiko von Pestiziden in der Tat überschätzt werden. Allerdings kann der Realismus in der Risikobewertung durch eine derartige Verfeinerung nur erhöht werden, wenn darin gleichermaßen weitere Prozesse berücksichtigt werden, welche das tatsächliche Potenzial zur Wiedererholung einer Population senken können. Leider steht die detaillierte Darstellung von Wiedererholungsprozessen in den Modellen im Widerspruch zu der geringen Aufmerksamkeit, welche solchen limitierenden Prozessen gewidmet wurde.

Bspw. spiegelte der Umfang der simulierten Eingangseffekte von Pestiziden auf Organismenebene in den meisten Fällen das tatsächliche Risiko nicht ausreichend wider. Die im Teil 3 dieses Berichts evaluierten Studien zur Wiedererholung einer Population nach einem akuten Populationseinbruch simulierten lediglich letale, aber keine subletalen Effekte. Allerdings führt eine Exposition, welche akute Mortalität verursacht, i. d. Regel zu chronischen Effekten in den überlebenden Individuen, wie z. B. einer Verringerung des Wachstums, der Reproduktion, der Konkurrenzstärke sowie der Fähigkeit, Räubern zu entkommen. Eine derartige Reduktion der Fitness über einen längeren Zeitraum kann die Wiedererholung einer Population erheblich erschweren (Desneux et al. 2006). Darüber hinaus können zusätzliche Stressoren im Freiland wie Nahrungsmangel oder ungünstige Temperaturen die Sensitivität von Individuen gegenüber den simulierten Effekten von Pestiziden erhöhen (Liess et al. 2016a). Toxizitätsmodule, welche derartige Interaktionen von Eingangseffekten auf Individuen-Ebene mit der Umwelt berücksichtigen sind selten, und Module, welche dazu grundsätzlich in der Lage sind (z. B. DEB), sind noch nicht ausreichend validiert.

Schließlich können Umweltstressoren wie ungünstige Temperaturen und Sauerstoffmangel in Gewässern oder räuberische und konkurrierende Arten das Populationswachstum und damit das Potential zur Wiedererholung beeinträchtigen (Liess et al. 2013). Wie Gabsi et al. (2014d) und Kattwinkel und Liess (2014) gezeigt haben, kann das Hinzufügen von antagonistischen Arten in Populationsmodellen die vorhergesagte Wiedererholung gravierend verändern. Allerdings wurden in keiner der in Teil 3 dieses Berichts evaluierten Studien antagonistische Arten explizit simuliert, und abiotische Stressoren wurden nur in einem geringen Ausmaß berücksichtigt. Dies ist vor allem dann problematisch, wenn physiologische Prozesse mit Labordaten unter vergleichbar günstigen Umweltbedingungen parametrisiert worden sind. Im Gegensatz dazu können Umweltstressoren zu einem gewissen Grad implizit in einem Model enthalten sein, wenn die physiologische Parametrisierung auf (Semi-)Freilanddaten beruht. Allerdings können Effekte selbst dann unterschätzt werden, da Individuen durch chronische Effekte von Pestiziden mit potenziell verringerter Fitness (s. o.) anfälliger gegenüber weiteren Umweltstressoren werden können (Becker und Liess 2015).

Insgesamt wurde daher eine Tendenz beobachtet, bevorzugt solche ökologischen Prozesse zu berücksichtigen, die das Risiko von Pestiziden senken können. Dadurch sind die Ergebnisse von Populationsmodellen wahrscheinlich verzerrt können und zu einer Unterschätzung des tatsächlichen Risikos führen.

Lebensgemeinschafts- und Ökosystemmodelle

Die allgemeinen Schutzziele für Lebensgemeinschaften und Ökosysteme fordern den Ausschluss inakzeptabler Effekte auf die Biodiversität und Ökosystemfunktionen, spezifische Schutzziele wurden bislang jedoch nicht eingeführt (Verordnug (EG) Nr. 546/2011). Das Fehlen von SPG für Lebensgemeinschaften und Ökosysteme sowie die hohe Komplexität der Parametrisierung und Validierung von Modellen stellen wesentliche Hindernisse für eine höherstufige Effektmodellierung in der Umweltrisikobewertung von Pestiziden dar. Entsprechend wurden keine Anwendungen von Lebensgemeinschafts- und Ökosystemmodellen in der prospektiven regulatorischen Risikobewertung gefunden.

Das Ökosystemmodell AQUATOX wurde explizit dazu entworfen, Effekte von zusätzlichen Stressoren auf die Wiedererholung einer Population sowie indirekte Effekte zwischen trophischen Ebenen und Biomagnifikation innerhalb eines Nahrungsnetzes zu untersuchen. Das Toxizitätsmodul für direkte Eingangseffekte kann akute und chronische Effekte berücksichtigen, aber keine Interaktionen mit zusätzlichen Stressoren (mehrere Stressoren wirken auf Organismus-Ebene additiv). Das Toxizitätsmodul wurde so konzipiert, dass lediglich minimale ökotoxikologische Informationen aus Tier 1-Standardtests benötigt werden. Anders wäre die Parametrisierung eines gesamten Ökosystemmodells mit zahlreichen Arten vermutlich nicht zu bewerkstelligen Allerdings basiert das Toxizitätsmodul deshalb auf zahlreichen Annahmen wie der Haber-Regel, die nicht in jedem Fall zutreffen und für erhebliche

Unsicherheiten in dem Modell sorgen können. Dies ist vermutlich der Hauptgrund dafür, dass AQUATOX und andere Ökosystemmodelle üblicherweise für die retrospektive Analyse von beobachteten Freilandeffekten verwendet werden, nicht aber für eine prospektive Risikobewertung.

Der $SPEAR_{pesticides}$-Ansatz wurde ursprünglich als Bioindikator entwickelt, um die Pestizidbelastung in kleinen Fließgewässern anhand charakteristischer Änderungen in der Makroinvertebraten-Lebensgemeinschaft einzuschätzen (Liess und von der Ohe 2005). $SPEAR_{pesticides}$ stellt daher ein Werkzeug für das Monitoring, aber auch für die Vorhersage von Pestizid-Effekten im Freiland dar. Der $SPEAR_{pesticides}$-Ansatz kann zum einen retrospektiv für eine Überprüfung der tatsächlichen Schutzwirkung der etablierten Methoden in der Umweltrisikobewertung eingesetzt werden kann. Der Ansatz kann aber auch prospektiv genutzt werden, um anhand einer empirisch etablierten Beziehung von Effekten auf Individuen-Ebene im Labor (in Form von toxic units, TU) auf Gemeinschaftseffekte (in Form von SPEAR-Werten) im Freiland zu schließen. SPEAR-Werte beschreiben das Verhältnis der Individuenzahlen von vulnerablen zu nicht-vulnerablen Taxa. Der SPEAR-Wert kann daher nicht direkt auf die spezifischen Schutzziele der EFSA angewandt werden, besitzt aber eine hohe Relevanz, um unmittelbar das tatsächliche Schutzziel einer hohen Sicherheit gegenüber Langzeitfolgen für die Abundanz und Diversität von aquatischen Invertebraten zu addressieren (Verordnung (EG) Nr. 1107/2009, EFSA PPR 2013). Der $SPEAR_{pesticides}$-Ansatz kann nicht auf die Bewertung eines spezifischen Pestizids mit einem bestimmten Anwendungsmuster und Wirkmechanismus angepasst werden. $SPEAR_{pesticides}$ bezieht aber Effekte auf die maximale TU, welche sich für ein spezifisches Pestizid anhand einer Fate-Modellierung vorhersagen lässt. Daher kann $SPEAR_{pesticides}$ für ein Screening nach Freilandeffekten eingesetzt werden, welche von einem Pestizid mit einer gegebenen Toxizität unter realistischen Bedingungen üblicherweise zu erwarten sind.; die vorhergesagten Effekte können als ein Bezugspunkt für potentiell entlastende fallspezifische Studien dienen.

Potenzial für Verbesserungen

Die Dokumentation der Entwicklung und Anwendung von Effektmodellen für die Risikobewertung war im Allgemeinen gut und folgte dem ODD Protokoll (overview, design concepts, detail; Grimm et al. 2006, Grimm et al. 2010) oder den TRACE Richtlinien, welche für die Beschreibung von Modellen vorgeschlagen worden sind (Grimm et al. 2014). Allerdings legt das EFSA Gutachten zur guten Modellierungs-Praxis (2014b) großen Wert auf die Evaluation eines Modells mithilfe von Sensitivitäts- und Unsicherheitsanalysen sowie durch Validierung, bevor es in der regulatorischen Risikobewertung eingesetzt werden kann. Für TKTD-Modelle ist eine Sensitivitätsanalyse nur von begrenztem Wert, da diese Modelle vollständig mittels spezifischer Daten kalibriert werden. Daher liefert in diesen Fällen die Anpassungsgüte der Kalibrierung vermutlich wichtige Informationen. Demgegenüber sind Sensitivitätsanalysen ein wichtiges Instrument für die Evaluation von Populationsmodellen und wurden auch üblicherweise für ein Standard-Szenario zur Entwicklung und Präsentation von Modellen für die wissenschaftliche Gemeinschaft zur Verfügung gestellt. Allerdings wurden die aus regulatorischer Sicht wichtigsten Endpunkte, d. h. die Sensitivität der Dauer bis zur Wiedererholung sowie die Sensitivität der NOAEL gegenüber verschiedenen Eingangsparametern (insbesondere der Stärke von Effekten auf Individuen-Ebene) in keinem der im Detail evaluierten Modelle berücksichtigt.

Daneben sollte die Validierung von Modellen erheblich verbessert werden. Einige der Populationsmodelle wurden bereits in der Risikobewertung angewendet, obwohl ihre Vorhersagen noch nie mit unabhängigen realen Daten überprüft worden waren. In anderen Fällen wurden lediglich Modellvorhersagen für die Populationsdynamik in Kontrollszenarien ohne Pestizideinsatz getestet, nicht jedoch die regulatorisch relevanten Endpunkte (Wiedererholungszeit oder NOAEL). Vorhersagen zur Dynamik in kontaminierten Populationen wurden lediglich in einem einzigen Fallbeispiel mit „realen“ Beobachtungen (experimentellen Daten aus einer Mesokosmenstudie) verglichen; in dieser experimentellen Studie wurden allerdings keine Langzeiteffekte beobachtet, so dass die Fähigkeit des Modells, tatsäch-

lich existierende unakzeptable Effekte vorherzusagen, nicht überprüft werden konnte. Der im Allgemeinen ungenügende Umfang der Validierung lässt sich vermutlich auf einen Mangel an geeigneten Daten zurückzuführen, auf welche ein Populationsmodell angewandt werden könnte. Allerdings sind die ökologischen Mechanismen in diesen Modellen, abgesehen vom Toxizitätsmodul für Eingangseffekte, unabhängig von der Art und Wirkungsweise eines bestimmten Schadstoffs. Daher bietet sich zur Validierung auch die Nutzung von historischen Daten für eine Vielzahl von Schadstoffen an. Für einen geeigneten Vergleich (Validierung) zwischen simulierten und beobachteten Daten empfiehlt es sich, die Zunahme von simulierten Populationseffekten (Ausgangsvariable) mit der Zunahme von Effekten auf Individuen-Ebene (Eingangsvariable) zu quantifizieren. Diese Beziehung kann anschließend mit der real beobachteten Beziehung zwischen Effekten auf Individuen- und Populationsebene verglichen werden.

Insgesamt gesehen können mechanistische Effektmodelle helfen, Effekte unter künstlichen Testbedingungen auf Situationen im Freiland zu übertragen; ihr Einsatz kann daher potenziell zu einem erhöhten Realismus in der Umweltrisikobewertung von Pestiziden führen. Allerdings sollte beim Design der Modelle auf ein ausgeglichenes Verhältnis von Prozessen geachtet werden, welche das tatsächliche Risiko potenziell vergrößern oder verringern, und die Evaluierung von Modellen sollte verbessert werden. Schlussendlich können Modelle lediglich helfen, das Risiko von solchen Effekten einzuschätzen, die *a priori* bekannt sind; ein potentielles Risiko durch neuartige Wirkmechanismen kann mit ihnen nicht erkannt werden. Daher können Effektmodelle experimentelle Arbeiten und Feldbeobachtungen unterschützen, sie aber nicht ersetzen.

1 Overview on Potentially Relevant Models for Effect Modelling

1.1 Introduction

In part 1 of this report, we briefly reviewed a number of effect models that may be of potential interest for the risk assessment of pesticides. The models range across various levels of biological organization from the individual to the ecosystem. Due to the large number of available models, this project reviews only a selection (and not a full overview) of models published until the year 2017. Models have been sorted by the addressed biological organization level, and by the principal conceptual modelling approach. These conceptual approaches are represented by one or several case studies. Each model review starts with a short introduction in continuous text format, followed by textboxes in tabular format on general properties, variables and parameters, evaluation and documentation, an assessment, and a list of important publications.

Reviewed models for the organization level of **organisms** (individuals) include some toxicokinetic (TK) models. Pure TK models actually model pesticide uptake and elimination rather than effects. However, they build the basis for the integrated toxicokinetic-toxicodynamic (TKTD) models for internal fate and effects, and therefore several variants have been addressed here. Reviewed TKTD models include the GUTS framework for damage-based modelling of acute mortality, and different forms of energy budget models for acute and chronic lethal and sublethal effects (DEB, NPM). In addition to these dynamic simulation models, static models have been established in ecotoxicology mainly to predict interacting effects of different pesticides with similar or different modes of action (variants of Concentration Addition CA and Effect Addition EA). These models have not been reviewed here. However, we reviewed a recently published static model for the prediction of effect interaction from pesticides and additional environmental stressors other than pesticides (Stress Addition Model, SAM).

Table 1: Individual-Level Models Briefly Evaluated

Model (Authors)	Main reviewed publication
Toxicokinetic (TK) Models	
1-Compartment Toxicokinetic (TK) Models	Spacie and Hamelink (1982)
2-Compartments Toxicokinetic (TK) Models	Spacie and Hamelink (1982)
Physiologically-based Toxicokinetic (PBTK) Models for Animals	Krishnan and Peyret (2009)
Physiologically-based Toxicokinetic (PBTK) Models for Plants	Trapp and McFarlane (1995)
Dynamic Energy Budget (DEB) as TK Model	Kooijman (2010)
Toxicokinetic-Toxicodynamic (TKTD) Models	
General Unified Threshold Model of Survival (GUTS)	Jager et al. (2011)
Dynamic Energy Budget (DEB) as TKTD Model	Kooijman (2010)
DEBtox	Jager and Zimmer (2012)
DEBkiss	Jager (2018)
Net-Production Models (NPM)	Brett and Groves (1979)
Static Models	
Stress Addition Model (SAM)	Liess et al. (2016b)

The reviewed **population** models include some discrete models that proceed in fixed time steps based on difference equations or matrix algebra (often preferred for demographic models that consider populations structured in cohorts of age classes or life stages). Additionally, we reviewed examples of continuous models based on ordinary differential equations (ODE) or on partial (PDE) and delayed (DDE) differential equations (for structured populations). Finally, we reviewed a number of the more recently developed individual-based models (IBMs) for risk assessment, in which each individual (or groups of individuals termed "superindividual") may have unique values for parameters and state variables. They are mainly used to capture temporal variability in the development of modelled organisms that may increase realism in predictions on pesticide effects and recovery at the population level. Spatially explicit IBMs additionally introduce complexity from spatial heterogeneity and are based on behavioral rules for the movement and resource utilization of individuals in a landscape; they have been mainly used in risk assessment to consider recolonization processes and spatial variability in exposure (refined proportion of feeding in treated sites, PT). All these model types can be coupled to modules (submodels) for the calculation of individual-level effects with varying complexity, ranging from simple dose-response models to the energy budget approaches outlined above. We reviewed examples for various combinations of population modelling approaches with individual-level modelling approaches. Additionally, we reviewed an example of a habitat suitability model. These static models are empirical and not mechanistic, but are widely used in conservation biology, and the example was explicitly designed for the prediction of pesticide exposure and effects in a heterogeneous landscape.

Table 2: Population Models Briefly Evaluated

Model (Authors)	Main reviewed publication
Discrete Models	
Simple Discrete Models	Calow et al. (1997)
Probabilistic Discrete Models	Fabre et al. (2006)
Discrete Models with Multiple Stages	Gledhill and Van Kirk (2011)
Simple Matrix Models	Charles et al. (2004)
Matrix Modelling with DEB	Klanjscek et al. (2006)
Matrix Modelling with DEBtox	Billoir et al. (2007)
Matrix Modelling with kmDEB	Klok and De Roos (1996)
Continuous Models	
Simple Ordinary Differential Equations (ODE) Models	Barnthouse (2004)
ODE Model Coupled to Population Size	Hendriks and Enserink (1996)
ODE Model for Aphids	Adams et al. (2005)
Staged ODE Model for Mosquitofish	Cabral et al. (2001)
Spatial ODE Models	Byers and Castle (2005)
Growth Model for Aquat. Plants with 1-Comp. TK	Schmitt et al. (2013)
Growth Model for Aquat. Plants with 3-Comp. TK	Heine et al. (2014)
DEB for Unicellulars	Hanegraaf and Muller (2001)
DEBtox for Unicellulars	Kooijman et al. (1996)
Euler-Lotka Equation with DEBtox	Jager et al. (2004)
Euler-Lotka Equation with kmDEB	Kooijman and Metz (1984b)
Partial Differential Equation (PDE) Model with Energy Budget	Hallam et al. (1990a)

Model (Authors)	Main reviewed publication
Delay-Differential Equation (DDE) Models	Brown et al. (2003)
Individual-Based Models (IBMs)	
Connected Individual and Population Models for Seals	Hickie et al. (2005)
BEEHAVE	Becher et al. (2014)
IDamP	Preuss et al. (2009a)
Chaoborus IBM Population Model	Strauss et al. (2016)
IBM with DEB	Martin et al. (2012)
IBM and PDE with kmDEB	Baveco and De Roos (1996)
IBM with NPM for D. magna	Vanoverbeke (2008)
Spatially Explicit IBMs	
Spatial IBM for Marine Crustaceans	de los Santos et al. (2015)
MASTEP	Van den Brink et al. (2007a)
IBM with GUTS for Aquatic Invertebrates	Baveco et al. (2014)
SpringSim	Meli et al. (2013)
Spatial IBM with Energy Budget for Earthworms	Johnston et al. (2014b)
eVole	Wang (2013)
IBM with TK for the Wood Mouse	Liu et al. (2013)
ALMaSS	Topping et al. (2003)
Empirical Models	
Habitat Suitability Models (HSM)	Chow et al. (2005)

Community or **food web** models address interactions between populations of different species that are connected via a food web. In contrast to ecosystem models, modelling of the abiotic environment only sets the stage for the survival and development of the populations. As a consequence, in community models, organisms are typically affected by the abiotic environment, but do not affect it themselves (apart from depleting food sources which are, however, replenished by external driving functions). We reviewed an example of a simple discrete model for parasite-host interactions, several continuous models that simulate increase and decrease in the overall biomass of populations using differential equations, and a spatial and a non-spatial IBM for multiple species. The bioaccumulation model of Arnot and Gobas (2004) is actually an exposure model, but shares many characteristics with other community effect models and could be easily linked to an external toxicity module for individual- (or population-) level effects. The model shares many features with an ecosystem model, but lacks detail in the description of the abiotic environment, so that we considered it as a community level model. For comparison, we additionally reviewed two empirical rather than mechanistic models that are available for the prediction of community level effects (PERPEST and $SPEAR_{pesticides}$).

Table 3: Community / Food Web Models Briefly Evaluated

Model (Authors)	Main reviewed publication
Discrete Models	
Model for Parasite-Host Interactions	Waage et al. (1985)
Continuous Models	
TK Model for Aquatic Bioaccumulation	Arnot and Gobas (2004)
ODE Model for Freshwater Communities	De Laender et al. (2007)
ODE Model for Pesticide Resistance	Becker and Liess (2015)
Individual Based Models (IBMs)	
IBM for Effects of Competition and Pesticides	Kattwinkel and Liess (2014)
Eco-SpaCE	Loos et al. (2010)
Empirical Models	
PERPEST	Van den Brink et al. (2002)
$SPEAR_{pesticides}$	Liess and von der Ohe (2005)

Ecosystem models focus on major processes in both the biotic (populations) and abiotic compartments (water column, sediment, etc.) of a whole ecosystem. E. g., ecosystem models typically simulate the cycling of nutrients and toxicants through various biotic and abiotic compartments. Organisms often can affect abiotic conditions in the models, e. g. photosynthesis and respiration may change pH and the amount of dissolved oxygen in a water body. All the reviewed models use differential equations to simulate fluxes of (bio)mass between the different compartments and keep mass balance in the modelled system (but allow for in- and outflow). All the models are not spatially explicit (though AQUATOX can simulate several connected segments of a water body) and, except for CATS, are limited to (natural or artificial) aquatic ecosystems.

Table 4: Ecosystem Models Briefly Evaluated

Model (Authors)	Main reviewed publication
Freshwater Models	
AQUATOX	Park et al. (2008)
CASM	Bartell et al. (1999)
CATS	Traas and Aldenberg (1996)
Streambugs	Schuwirth and Reichert (2013)
Chemostat Model with DEB	Kooi et al. (2008b)
Saltwater Models	
ECOWIN	Ferreira (1995)
NEMURO	Kishi et al. (2007)

1.2 Individual Level Models

1.2.1 Toxicokinetic (TK) Models

1.2.1.1 1-Compartment TK Models (Spacie and Hamelink 1982)

The one-compartment model is the basic model for toxicokinetics (TK). It is based on the assumptions that the organism can be represented by one well-mixed compartment, and that uptake is proportional to the external concentration and elimination proportional to the internal concentration (in its simplest form). In this form, the model is used to analyse body-residue data, and forms a building block for toxicokinetic-toxicodynamic (TKTD) models. Spacie and Hamelink (1982) provided an early review on many extensions to account for growth, uptake from feeding, saturating kinetics, etc.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Can be used at Tier 1 in ERA.
Model purpose	Scientific / Regulatory	The original intention of the model is unclear. Over the last decade it has been widely applied in both scientific and regulatory contexts.
Questions / processes	Body burden	Model aim is to explain body burdens of individuals over time. The basic model has been extended in many ways.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Due to the assumption of rapid internal redistribution, not always suitable for large organisms (e.g., birds and mammals) and “slow” chemicals.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Applied for registration	The model is well-established in science and already applied for registration. It is included in OECD guidelines for bioconcentration tests.
Public availability	Software extension	Countless implementations exist, probably also as stand-alone application. The basic model can be implemented in any statistical software (and also in Excel).

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	The organism is represented as a single well-mixed compartment.
Endpoints	Body residues BCF	The model can be fitted to body-residue data and thereby provide bioconcentration factor (BCF), elimination and uptake rate constants. It can predict body residues as a result of time-varying exposure.

Criteria	Categories	Comments
Space	No spatial context	
Time	Dates	Generally days – weeks – months.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Can also be applied to mixtures (generally assuming that compounds do not interact).
Abiotic environment	Temperature	The abiotic environment is only represented by the chemical concentration. Work has been done on temperature-dependence of model parameters, and on the contribution from feeding.
Biotic environment	None	None.
Individuals	Homogeneous	Individual represented as a single well-mixed compartment.
Populations	-	TK models deal with individuals, but can be linked to TKTD and population approaches.
Calibration	Laboratory data Mesocosm data Field data	TK models are fitted to experimental data, and thus calibrated for each case.
Programming language	Many	The model (in its simplest forms) can easily be implemented in any statistical software (as well as in Excel).

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Mesocosm data Field data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication	Concepts and equations are explained in almost every textbook on ecotoxicology and environmental risk assessment.

Assessment

Criteria	Description
Strengths	Can be calibrated using standard laboratory tests; simple model; useful to describe body residues in many cases, and allows extrapolation to other exposure scenarios.
Theoretical uncertainties	Based on a rigorous simplification of animals into a single well-mixed compartment. Therefore, many details are lost.
Empirical uncertainties	Many species or data sets require adaptations to the basic model; for example, to deal with growth, metabolism, or saturating kinetics.
Parametric uncertainties	
Temporal uncertainties	Parameters generally assumed to remain constant over time
Conclusions	The one-compartment TK model is an indispensable and integral part of TKTD modelling and ERA. It is simple, and provides a good explanation of body residues over time in many cases.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Widmark and Tandberg (1924)	Mammals?	Narcotics	The first publication we found that deals with the one-comp. model in a TK context.
Spacie and Hamelink (1982)	Fish	General	One of the early ecotoxicology reviews mentioning various extensions of the one-comp. model.
Model applications			
De Bruijn and Hermens (1991)	Guppy	Organophosphorous pesticides	No simultaneous fitting of uptake and elimination phase.
Gobas et al. (1986)	Fish (general)	General	Parameters are related to underlying physiology of the fish and hydrophobicity of the chemicals.
Barber et al. (1988)	Fish (general)	Nonpolar organics	Parameters related to underlying physiology, includes dilution by growth.
Hendriks et al. (2001)	Large range of plants and animals	General	Parameters correlated to underlying physiology of the organism and hydrophobicity of the chemicals. Uptake from various routes (incl. food), elimination through different processes, and growth dilution (assuming exponential growth).
Kooijman and Bedaux (1996b)	General	General	General extension for growth with dilution (for any type of growth) and changes in surface-volume ratio.

Citations	Taxa	Chemicals	Comments
Jager et al. (2003)	Earthworms	Chlorobenzenes and PCB153	Extension with a first-order 'disappearance' rate from soil to reflect changes in bioavailability, and a dynamic compartment for the gut contents.
Jager et al. (2000)	Earthworms	PAHs	Extension with a first-order 'disappearance' rate from soil to reflect changes in bioavailability.
Rubach et al. (2010)	15 freshwater arthropod species	Chlorpyrifos	Uptake and elimination fitted, using measure water concentrations over time. Conf. intervals on both model parameters and model curve.
Bednarska et al. (2013)	Rat	Thiamethoxam	Standard model to fit concentrations in blood over time. Extension with a simple gut compartment to simulate whole-body concentrations over time under realistic feeding regimes.

1.2.1.2 2-Compartments TK Models (Spacie and Hamelink 1982)

The two-compartments model is an obvious extension of the one-comp. model for TK. It is based on the assumptions that the organism can be represented by two well-mixed compartments, and that uptake is proportional to the external concentration and elimination proportional to the internal concentration in each compartment (in its simplest form). The two compartments can represent two parts of the body (e.g., structure and lipid storage) or two forms of the chemical (e.g., parent and metabolite). The model is used to analyze body-residue data, and forms a building block for TKTD models. Spacie and Hamelink (1982) provided an early review on many extensions that have been presented to account for growth, uptake from feeding, saturating kinetics, etc.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Can be used at Tier 1 in ERA.
Model purpose	Scientific / Regulatory	Unclear what the original intention of the model was. Over the last decade it has been widely applied in both scientific and regulatory contexts.
Questions / processes	Body burden	Model aim is to explain body burdens of individuals over time. The basic model has been extended in many ways.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Due to the assumption of only two internal compartments, not always suitable for large organisms (e.g., birds and mammals) and 'slow' chemicals.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	The two-comp. model is well-established in science, but less used in a regulatory context.
Public availability	Software extension	Countless implementations exist, probably also as stand-alone application. The most basic model can be implemented in any statistical software (and also in Excel).

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	The organism is represented as two well-mixed compartments.
Endpoints	Body residues BCF	The model can be fitted to body-residue data and thereby provide BCF, elimination and uptake rate constants. It can predict body residues as a result of time-varying exposure.
Space	No spatial context	
Time	Dates	Generally days – weeks – months.

Criteria	Categories	Comments
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Can also be applied to mixtures (generally assuming that compounds do not interact).
Abiotic environment	Temperature	The abiotic environment is only represented by the chemical concentration. Work has been done on the contribution from feeding, and possibly also on temperature dependence.
Biotic environment	None	None.
Individuals	Homogeneous	Individual represented as two well-mixed compartments.
Populations	-	TK models deal with individuals, but can be linked to TKTD and population approaches.
Calibration	Laboratory data Mesocosm data Field data	TK models are fitted to experimental data, and thus calibrated for each case.
Programming language	Many	The model (in its simplest forms) can easily be implemented in any statistical software (as well as in Excel).

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Mesocosm data Field data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication	Concepts and equations are explained in many textbooks on ecotoxicology and environmental risk assessment.

Assessment

Criteria	Description
Strengths	Can be calibrated using standard laboratory tests; simple model; useful to describe body residues in many cases, and allows extrapolation to other exposure scenarios.
Theoretical uncertainties	Based on a rigorous simplification of animals into two well-mixed compartments. Therefore, many details are lost.
Empirical uncertainties	Many species, or data sets, require adaptations to the basic model; for example, to deal with growth, metabolism, or saturating kinetics.
Parametric uncertainties	
Temporal uncertainties	Parameters generally assumed to remain constant over time.

Criteria	Description
Conclusions	The two-comp. TK model is a logical extension of the one-comp. model. It is simple, and in some cases provides a better explanation of body residues over time.

Publication

Citations	Taxa	Chemicals	Comments
		Model description	
Könemann and Van Leeuwen (1980)	Guppy	Chlorobenzenes	One of the early applications of the two-comp. model in ecotox.
Spacie and Hamelink (1982)	Fish	General	One of the early ecotox reviews mentioning the two-comp. model.
		Model applications	
Steen Redeker et al. (2004)	Tubifex	Cadmium and zinc	Basically, a three-comp. model as it includes two compartments for the organism and a gut compartment.
Jager et al. (2003)	Earthworms	Chlorobenzenes and PCB153	Extension with a first-order 'disappearance' rate from soil to reflect changes in bioavailability, and a dynamic compartment for the gut contents.
Bednarska et al. (2013)	Rat	Thiamethoxam	Standard model to fit concentrations in blood over time. Extension with a simple gut compartment to simulate whole-body concentrations over time under realistic feeding regimes.
Van Eijkeren et al. (2006)	Chicken	Dioxins and PCBs	Chicken modelled as two compartments (central and fat compartment) to estimate concentrations in eggs.
Ducrot et al. (2015)	Skylark and woodmouse	Hypothetical	Application of a two-comp model (Bednarska et al) in context of ERA for PPPs.
Spann et al. (2015)	Nematodes	Phenanthrene	Central and peripheral compartment.

1.2.1.3 Physiologically-Based Toxicokinetic (PBTK) Models for Animals (Krishnan and Peyret 2009)

PBTK models consist of a series of well-mixed compartments (representing organs or tissue groups) linked by a blood flow. It describes how chemicals are taken up (e.g., from the gut) and distributed over the body. In some cases, metabolites are included as well. Krishnan and Peyret (2009) provided an overview of PBTK modelling concepts and applications in ecotoxicology.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Generally for Higher Tier ERA.
Model purpose	Scientific / Regulatory	First developed for human pharmacology (both scientific and for regulatory purposes).
Questions / processes	Body burden	Model aim is to explain body burdens of individuals over time. It extends the one- and two-comp. models by including a compartment for various organs (or groups thereof) and a blood flow between them.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Due to the inclusion of blood flow and (groups of) organs, specifically suited for vertebrates and 'slow' chemicals.
Taxon specificity	Generic	Mainly vertebrates.
Toxicant specificity	Generic	
Application	Applied for registration	The model is well-established in science and already applied for registration (specifically in the context of human health).
Public availability	-	

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	The organism is represented by a series of well-mixed compartments, linked by a blood flow.
Endpoints		The model can be fitted to body-residue data, and can predict body residues as a result of time-varying exposure. It is also used to extrapolate data from laboratory animals to humans.
Space	No spatial context	
Time	Dates	Generally weeks – months – years.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Can also be applied to mixtures (generally assuming that compounds do not interact).

Criteria	Categories	Comments
Abiotic environment		The abiotic environment is only represented by the chemical concentration.
Biotic environment	-	None.
Individuals		Individual represented as a set of well-mixed compartments.
Populations	-	TK models deal with individuals, but can be linked to TKTD and population approaches.
Calibration	Laboratory data	TK models are fitted to experimental data, or parameterized from animal tests and physico-chemical data.
Programming language	Various	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	General overview of validation status in (Chipps and Wahl 2008). They conclude that agreement between predictions and data is generally poor, but the conceptual basis of the models is valid.
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication	Concepts and equations are explained in many textbooks on pharmacology and toxicology.

Assessment

Criteria	Description
Strengths	Can include many details of the individual, and predict concentration profiles in different tissues.
Theoretical uncertainties	For cases where internal redistribution is fast compared to the exchange with the outside world, a PBTK model is overkill, and a one-comp. model will perform equally well.
Empirical uncertainties	Calibrations are species specific and rather intensive, but basic model parameters (like blood flow and organ weights) only need to be established once for a species.
Parametric uncertainties	
Temporal uncertainties	Parameters generally assumed to remain constant over time.
Conclusions	PBTK models are well established in pharmacology and human health risk assessment. They are mainly used for vertebrates.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Jacquez et al. (1960)	Humans	Pharmaceuticals	Possibly the first published PBTK model.
Krishnan and Peyret (2009)	Various vertebrates	Various	Overview of PBTK modelling concepts and applications in ecotoxicology.
Model applications			
Nichols et al. (1990)	Rainbow trout	Pentachloroethane	General PBTK model for fish.
Lien et al. (1994)	Fathead minnow	Chlorinated ethanes	Only three tissue compartments used.
Stadnicka et al. (2012)	Rainbow trout and fathead minnow	39 chemicals	Validation of model predictions for various chemicals, and comparison to the one-compartment model.
USEPA (2006)	Various mammals (relates to human health RA)	Generic	Overview of applications of PBTK models for human health RA.

1.2.1.4 Physiologically-Based Toxicokinetic (PBTK) Models for Plants (Trapp and McFarlane 1995)

These PBTK models for plants consist of a series of well-mixed compartments (representing plant parts such as roots, stems, leaves and fruit) linked by a translocation flow in the xylem and phloem. It describes how chemicals are taken up (by roots and leaves, i.e. from soil and air) and distributed over the tissues. In some cases, metabolites are included as well. The textbook of Trapp and McFarlane (1995) deals with plant PBTK modelling in detail.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Generally Higher Tier, although using QSARs to estimate toxicant-specific parameters could make them applicable to Tier 1 ERA.
Model purpose	Scientific / Regulatory	Developed with regulatory purpose in mind, though mainly applied in a scientific setting.
Questions / processes	Body burden	The aim of modelling is to explain body burdens of individual plants over time. Compartments represent plant parts (e.g., roots, stems, leaves and fruit).
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Generally intended for higher plants (for plants like algae and duckweed, simpler models may be more appropriate).
Taxon specificity	Generic	
Toxicant specificity	Generic	Developed for neutral organics, but extended to deal with ionizing chemicals as well.
Application	Established in science	The model is well-established in science. We are unsure whether they have been actually used in RA, but, if so, they would be used foremost to predict residues in crops for human consumption (and wildlife).
Public availability	Software extension	Models are available at the website of Stefan Trapp in Excel format: http://homepage.env.dtu.dk/stt/

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	The individual plant is represented by a series of well-mixed compartments.
Endpoints	Reproduction	The model can be fitted to residue data in plant tissue, but chemical-specific parameters can also be estimated from phsyico-chemical properties.
Space	No spatial context	
Time	Dates	Generally weeks – months – years.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations (time or space)	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known.

Criteria	Categories	Comments
	Repeated exposure Toxicant mixtures	
Abiotic environment	Food limitation	The abiotic environment is only represented by the chemical concentration.
Biotic environment	None	None.
Individuals	-	Individual represented as a set of well-mixed compartments.
Populations	-	TK models deal with individuals, but can be linked to TKTD and population approaches.
Calibration	Laboratory data	TK models are generally fitted to experimental data, or parameterized from physico-chemical data using QSARs.
Programming language	Excel	The plant models of Trapp and co-workers are available in Excel format at: http://homepage.env.dtu.dk/stt/

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication	Concepts and equations are explained in a textbook and in several papers.

Assessment

Criteria	Description
Strengths	Can include many details of the plants, and predict concentration profiles in different tissues. Based on well-established principles. Simple enough to implement in Excel.
Theoretical uncertainties	Reducing plants to a few homogeneous compartments includes that details are lost (e.g., distribution within a leaf).
Empirical uncertainties	Calibrations are species specific and rather intensive, but basic model parameters (like translocation flow rates) only need to be established once for a species.
Parametric uncertainties	
Temporal uncertainties	Parameters generally assumed to remain constant over time.
Conclusions	These models are relatively simple, yet based on established principles. They have been applied for more than two decades. As such, they are prominent models in the field of TK for terrestrial plants.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Trapp et al. (1990)	Barley	Several, incl. atrazine and dieldrin	One of the first published plant models with several compartments.
Trapp and McFarlane (1995)	Various	Various	Textbook dealing with plant PBTK modelling in detail.
Model applications			
Trapp et al. (1994)	Soybean	Bromacil	Model description and validation experiments
Trapp (2000)	Soybean	Range of neutral and ionizable chemicals	Extension for ionizable compounds
Rein et al. (2011)	Various crop species	Hypothetical chemicals	Extensions to make the plant model more applicable to dynamic exposure situations
Legind et al. (2011)	Pepper	Methomyl	Concentrations in pepper fruits due to drip irrigation with an insecticide
Paterson et al. (1994)	Soybean	Four organic compounds (incl. bromacil)	Slightly different model, expressed in fugacity terms.
Fujisawa et al. (2002)	Japanese radish	Furametpyr, pyriproxyfen	Different model on similar principles, focusing on root crops.
Trapp and Eggen (2013)	Barley and carrot	Various, including 3 OPs	

1.2.1.5 Dynamic Energy Budget (DEB) as TK Model (Kooijman 2010)

Dynamic Energy Budget (DEB) is not a single model but a theoretical framework described in Kooijman (2010) from which various models can be derived (see section 1.2.2.2). The theory deals with metabolic organization; how resources are taken up from food and used to fuel energy-requiring processes (growth, reproduction, maintenance, etc.). This entry focusses on the application of DEB models applied only as toxicokinetic (TK) model.

General Properties

Criteria	Categories	Comments
Biological level	Individual	
Model purpose	Scientific	DEB is a general theory for metabolic organization, from which specific models can be derived for various purposes. This entry concerns the application as TK model (no toxicity).
Questions / processes	Body burden	DEB deals with development, growth and reproduction over the life cycle. These processes interact with the uptake and elimination of toxicants.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Not applied in regulatory applications.
Public availability	Software extension	Various implementations.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	Body residues in growing (and reproducing) organisms.
Endpoints		
Space	No spatial context	
Time	Dates	Generally days – years.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known.
Abiotic environment	Food limitation Temperature	Food is integral part of all energy-budget models. Temperature affects all rate constants using the Arrhenius relationship.
Biotic environment	None	DEB models focus on the individual level; biotic processes such as competition would need additional assumptions/modules.

Criteria	Categories	Comments
Individuals	Energy budget	Energy budget for growth and reproduction.
Populations	-	DEB deals with individuals, but can be (and has been) linked to population approaches.
Calibration	Laboratory data Field data	DEB models are generally fitted to experimental data, and thus calibrated for each case.
Programming language	Unclear	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Field data	
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication Website	DEB theory is described in detail in a book by Kooijman (2010). Concepts are also explained in a free e-book (Jager 2019), incl. the application to toxicant stress. Manuals are available for various software implementations.

Assessment

Criteria	Description
Strengths	Powerful framework to link many aspects of a species life history (growth, reproduction, respiration, body composition, product formation etc.). There is a large database with parameters available for a broad range of species, which is maintained by a group of 'curators'. There is a substantial community of people working on DEB, an intensive course is offered, and an international symposium.
Theoretical uncertainties	Based on a rigorous simplification of animal energetics over the entire life cycle. Inevitably, details on life history will be lost.
Empirical uncertainties	Many species, or data sets, require adaptations to the basic model; often difficult to find a unique mechanism of action of a chemical; no applications to toxic stress in birds and mammals.
Parametric uncertainties	Parameterization is complex, and usually requires additional assumptions or general rules for inter-species extrapolations.
Temporal uncertainties	Parameters generally assumed to remain constant over time.
Conclusions	DEB is a theory from which specific models can be derived. 'Full' DEB models are rather difficult to parameterise (in a unique way), but offer a powerful platform to capture many aspects of an organismsm life history in a single consistent framework. For most RA applications, simplifications such as DEBtox and DEBkiss are likely more useful.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman (2010)	Various	Various	This book explains the DEB theory in detail. Chapter 6 deals with toxicity and provides some examples.
Model applications			
Van Haren et al. (1994)	Mussels	Cd, PCBs, PAHs	
Klanjscek et al. (2007)	Whales		Modified DEB model, integrated with a pharmacokinetic module
Bodiguel et al. (2009)	Marine fish (hake)	PCBs	
Eichinger et al. (2010)	Marine fish (sole)	PCBs	
Noonburg et al. (2010)	Marine mammals	Hypothetical	Includes transfer of toxicants from mother to offspring

1.2.2 Toxicokinetic-Toxicodynamic (TKTD) Models

1.2.2.1 General Unified Threshold Model of Survival (GUTS, Jager et al. 2011)

GUTS (General Unified Threshold Model of Survival) links a one-compartment TK model, via a one-compartment damage model, to a death mechanism (stochastic death, individual tolerance, or a mixture of both). By fixing parameters to specific values, special cases can be derived, including virtually all of the TKTD survival models that have ever been used. Thus, GUTS is more a modelling framework than a single model and has been described in Jager et al. (2011) and Jager and Ashauer (2018b).

General Properties

Criteria	Categories	Comments
Biological level	Individual	Can be used at Tier 1.
Model purpose	Scientific / Regulatory	Several predecessors were specifically intended for regulatory purposes, but these models are more widely used for scientific purposes.
Questions / processes	Individual effects	Dose-response analysis for acute mortality/immobility data, and extrapolation to other exposure patterns.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Application to birds and mammals is rare, but not excluded (although it might require a more elaborate TK model).
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Sporadically applied in regulatory applications, included in OECD/ISO guidance.
Public availability	Software extension	Implementations in Matlab, R, Mathematica, ModelMaker. An executable also exists (Delphi).

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality	Can be used for other quantal endpoints such as immobility (although reversibility of the effect needs further study).
Endpoints	Survival	Survival of individuals. Statistics that can be calculated are LCx,t for any effect size x and time point t, or survival probabilities as a consequence of any exposure profile.
Space	No spatial context	
Time	Dates	Generally days – weeks.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known.

Criteria	Categories	Comments
	Toxicant mixtures	
Abiotic environment	Food limitation Temperature	GUTS is based on a stochastic representation of death in individuals. No consideration of abiotic factors, although temperature could be included as an effect on rate constants.
Biotic environment	None	No consideration of biotic factors, although predation or parasites could be added as additional death processes.
Individuals	Stochastic	The mortality process is viewed as stochastic at the level of the individual or of the cohort.
Populations	-	
Calibration	Laboratory data	GUTS models are fitted to experimental data, and thus calibrated for each case.
Programming language	Various	E.g., Matlab, R

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	GUTS models have been used to fit an enormous range of survival data sets. Some validation of the ability to extrapolate between exposure scenarios. See general explanation in report.
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication Website	The model equations are provided in the publication. No dedicated user manual for GUTS available at the moment (as far as we know), although many software implementations provide a manual of sorts. In 2017, a CEFIC-LRI funded project will produce an extensive guidance document (in the form of a free e-book) on GUTS.

Assessment

Criteria	Description
Strengths	Can be calibrated using standard acute laboratory tests; simple model; broad support; can use all information over time, and extrapolate to other exposure scenarios.
Theoretical uncertainties	Assumes a threshold for effects, and that death can be treated as stochastic (either at the individual level or on the group of individuals tested); unclear whether lab animals are relevant proxies for field populations. Unclear whether stochastic death or individual tolerance dominate as main mechanism of death (advisable to fit at least both extremes).
Empirical uncertainties	Estimating probabilities from observed frequency of response is difficult; accurate estimations require large numbers of individuals. Some data sets need additional mechanisms to get the model to fit.
Parametric uncertainties	Parameters will be different when assuming a different death mechanism; often difficult to identify all parameters with sufficient accuracy from standard data sets.

Criteria	Description
Temporal uncertainties	Parameters assumed to remain constant over time.
Conclusions	GUTS can directly be used to estimate LCx,t values from datasets (including tests where exposure varies), and thereby improves upon the standard dose-response analyses. In extrapolation to untested situations, additional uncertainties need to be considered.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Jager et al. (2011)	Amphipods and fathead minnow	Diazinon, naphthalene, trimethylbenz.	GUTS framework and illustration with several examples
http://www.debtox.info/about_guts.html			General web page
http://www.ecotoxmodels.org/guts/			General web page
Model applications			
Mackay et al. (1992)	Fathead minnow	Acetone	One of the first publications on the dynamic CBR model (the IT version in GUTS).
Bedaux and Kooijman (1994)	Guppy	Dieldrin, potassium dichromate	First published hazard model for survival.
Ashauer et al. (2007a)	Amphipod	Pentachlorophenol and chlorpyrifos	Extension with damage kinetics. Application to pulse exposures.
Jager and Kooijman (2005)	Fathead minnow	Various OP pesticides	Extension with a module for receptor kinetics.
Jager and Kooijman (2009)	Fathead minnow	All narcotics and reactives from minnow data base	Deriving relationships between parameters (and with Kow).
Baas et al. (2007)	Springtails	Heavy metals	General mixture approach. Application to binary mixtures.
Nyman et al. (2012)	Amphipod	Propiconazole	Analysis on various ways to calibrate the model, and validation of predictive power.
Ashauer et al. (2015)	Amphipod	14 compounds from different groups	Patterns in chemical space, detailed analysis of differences between SD and IT.
Ducrot et al. (2015)	Fathead minnow	Hypothetical	Case study using GUTS in a PPP ERA setting
Ashauer et al. (2016)	Various	Various pesticides	Validation of predictive abilities across exposure patterns, and link to dynamic SSDs
http://www.debtox.info/papers_survival.html			Full list of papers using hazard models (SD cases of GUTS)

1.2.2.2 Dynamic Energy Budget (DEB) as Toxicokinetik-Toxicodynamic Model (Kooijman 2010)

DEB, the Dynamic Energy Budget theory for metabolic organization (summarized in the textbook of Kooijman 2010), is not a single model but a theoretical framework from which various models can be derived. The theory deals with the question how resources are taken up from food and used to fuel energy-requiring processes (growth, reproduction, maintenance, etc.). DEB models are usually not based on molecular or physiological details of a species but rather focus on (rather abstract) lumped energy flows. This makes the theory generic for all living organisms. The DEBtox simplification follows from DEB theory but has received its own entry in this database.

General Properties

Criteria	Categories	Comments
Biological level	Individual	For dividing organisms, the distinction between individual and population becomes blurred.
Model purpose	Scientific	Intended as a general theory for metabolic organisation, from which specific models can be derived for various purposes. However, there is currently an EFSA-funded project running that looks at DEB as the link to population and mixture effects.
Questions / processes	Individual effects Effect propagation	Aim is to explain growth and reproduction (and survival) of individuals over time, including the effect of toxicants. Also linked to population-level calculations with Euler-Lotka equation, IBM and matrix models.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Applications for birds or mammals have been done, but not in a toxicological context.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Not applied in regulatory applications, although DEBtox is derived from the theory, which has been included in ISO/OECD guidance. An EFSA-funded project is currently running on application of DEB models in RA.
Public availability	Software extension	Matlab (within the framework of DEBtool: http://www.bio.vu.nl/thb/deb/index.html -> laboratory). Work is underway to develop an R implementation.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction	
Endpoints	Survival Reproduction	Statistics that can be calculated are LCx,t or ECx,t for any effect size x and time point t, or effects as a consequence of any exposure profile. Can be linked to population models.

Criteria	Categories	Comments
Space	No spatial context	
Time	Dates	Generally days – years.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Basic mixture effects have been added.
Abiotic environment	Food limitation Temperature	Food is an integral part of all energy-budget models. Temperature affects all rate constants using the Arrhenius relationship.
Biotic environment	None	DEB models focus on the individual level; biotic processes such as competition would need additional assumptions/modules.
Individuals	Stochastic Energy budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	-	DEB deals with individuals, but can be (and has been) linked to population approaches.
Calibration	Laboratory data	DEB models are generally fitted to experimental data, and thus calibrated for each case.
Programming language	Matlab	Several Matlab implementations are available; an R implementation is being made.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Website	DEB theory is described in detail in a book by Kooijman. Concepts are also explained in a free e-book (by Jager), incl. the application to toxicant stress. Manuals are available for various software implementations.

Assessment

Criteria	Description
Strengths	Powerful framework to link many aspects of a species life history (growth, reproduction, respiration, body composition, product formation etc.). There is a large database with parameters available for a broad range of species, which is maintained by a group of 'curators'. There is a substantial community of people working on DEB, an intensive course is offered, and an international symposium.
Theoretical uncertainties	Based on a rigorous simplification of animal energetics over the entire life cycle. Inevitably, details on life history will be lost.

Criteria	Description
Empirical uncertainties	Many species, or data sets, require adaptations to the basic model; often difficult to find a unique mechanism of action of a chemical; no applications to toxic stress in birds and mammals.
Parametric uncertainties	Parameterisation is complex, and usually requires additional assumptions or general rules for inter-species extrapolations.
Temporal uncertainties	Parameters generally assumed to remain constant over time.
Conclusions	DEB is a theory from which specific models can be derived. 'Full' DEB models are rather difficult to parameterise (in a unique way), but offer a powerful platform to capture many aspects of an organism life history in a single consistent framework. For most RA applications, simplifications such as DEBtox and DEBkiss are likely more useful.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Nisbet et al. (2000)	Various	Not specific for toxicity	General paper on DEB theory
Kooijman (2001)	Various	Not specific for toxicity	General paper on DEB theory
Kooijman (2010)	Various	Various	This book explains the DEB theory in detail. Chapter 6 deals with toxicity and provides some examples.
Jager et al. (2010)	Daphnia	Fluoranthene and pyrene (single and mixture)	Basic framework for chemical stress in the standard DEB model, with general extension to mixture effects.
http://www.bio.vu.nl/thb/deb/index.html			web page
http://ted.europa.eu/udl?uri=TED:NOTICE:162596-2015:TEXT:EN:HTML&tabId=1			Link to the EFSA tender
Model applications			
Augustine et al. (2012)	Zebrafish	Uranium	In-depth analysis.
Jager and Selck (2011)	Polychaete (Capitella)	Nonylphenol	Worked out case study
Martin et al. (2012)			General paper on the combination of DEB theory with IBM population modelling.
Martin et al. (2013b)	*Daphnia*	Dichloroaniline	Using DEB-IBM for a chemical stress. Validation with population data.

1.2.2.3 DEBtox (Jager and Zimmer 2012)

DEBtox is a simplified version of the standard DEB animal model, with the specific aim to deal with standard toxicity data in a regulatory context (fish growth and *Daphnia* reproduction). It has since been extended to life-cycle tests, simultaneously fitting growth, reproduction and survival over time. The latest derivation and statistical framework has been described in Jager and Zimmer (2012). Being a DEB-based model, DEBtox follows an energy budget where resources from food are allocated to energy-requiring processes (growth, reproduction, maintenance, etc.). Chemical uptake is covered by a one-compartment TK model, accounting for growth (no measured body residues are needed). Chemical stress affects one (or few) of these processes in the energy budget, leading to specific patterns of effects over the life cycle.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Can be used at Tier 1.
Model purpose	Scientific / Regulatory	Originally intended for regulatory purposes (to analyse standard toxicity data on growth and reproduction), but more widely used for scientific purposes.
Questions / processes	Individual effects Effect propagation	Model aim is to explain growth and reproduction (and survival) of individuals over time, including the effect of toxicants. Also linked to population-level calculations with Euler-Lotka equation or matrix models
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. No applications to birds or mammals yet (not clear whether it can be used/is useful for these groups).
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Sporadically applied in regulatory applications, included in OECD/ISO guidance. Mentioned by EFSA as promising tool for RA of bees (outsourced project ongoing).
Public availability	Software extension	Implementations in Matlab. An executable also exists, but is out of date (and will not be maintained).

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction	The model is a simplification following from Dynamic Energy Budget theory, using simple-to-interpret compound parameters instead of bioenergetic parameters.
Endpoints	Survival Reproduction	Statistics that can be calculated are LCx,t or ECx,t for any effect size x and time point t, or effects as a consequence of any exposure profile. Can be linked to population models.
Space	No spatial context	

Criteria	Categories	Comments
Time	Dates	Generally days – weeks – months.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Basic mixture effects can easily be added from related DEB-based models.
Abiotic environment	Food limitation Temperature	Food is integral part of all energy-budget models. Temperature affects all rate constants using the Arrhenius relationship.
Biotic environment	None	DEBtox models focus on the individual level; biotic processes such as competition would need additional assumptions/modules.
Individuals	Stochastic Energy budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	-	DEBtox deals with individuals, but can be linked to population approaches.
Calibration	Laboratory data	DEBtox models are fitted to experimental data, and thus calibrated for each case.
Programming language	Matlab	Several Matlab implementations are available. For example, within the BYOM framework: http://www.debtox.info/byom.html

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication Website	Concepts are explained in a free e-book (equations separate document), manuals are available for various software implementations.

Assessment

Criteria	Description
Strengths	Can be calibrated using laboratory tests (most useful: partial life-cycle tests); simple model; integrates information on different endpoints over time, and extrapolate to other exposure scenarios.
Theoretical uncertainties	Based on a rigorous simplification of animal energetics over the entire life cycle. Inevitably, details on life history will be lost.
Empirical uncertainties	Many species, or data sets, require adaptations to the basic model; often difficult to find a unique mechanism of action of a chemical; no applications to birds and mammals.

Criteria	Description
Parametric uncertainties	Often difficult to identify all parameters from standard data sets.
Temporal uncertainties	Parameters assumed to remain constant over time.
Conclusions	DEBtox can directly be used to estimate NEC or ECx values from datasets (including tests where exposure varies), and thereby improves upon the standard dose-response analyses. In extrapolation to untested situations, additional uncertainties need to be considered.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman and Bedaux (1996b)	*Daphnia magna*	Cadmium, phenol, dichloraniline	Focus on reproduction effects only, and on application to standard toxicity data sets.
Billoir et al. (2008b)	*Daphnia magna*	Copper and zinc	Correcting errors in the model derivation for some MoA's.
Jager and Zimmer (2012)	*Daphnia magna*	Fluoranthene	New derivation (correcting errors), incl. statistical framework.
http://www.debtox.info/about_debtox.html			General web page
Model applications			
Jager et al. (2004)	Springtails	Cadmium, tributyltin	Simultaneous fitting on all endpoints, ageing module amd population growth rate.
Pieters et al. (2006)	*Daphnia magna*	Fenvalerate	Model fitted to data for pulse exposed animals, at two food levels.
Alda Álvarez et al. (2006b)	Nematodes	Carbendazim and pentachlorobenzene	Model adaptation for nematodes, and demonstrating EC10 vs. time.
Muller et al. (2010)	*Daphnia*	tetradifon, pyridine	Slightly different DEB-based model, extrapolation to different food levels.
http://www.debtox.info/papers_debtox.html			Full list of papers using DEBtox models

1.2.2.4 **DEBkiss (Jager 2018)**

DEBkiss (described in Jager 2018) is a simplified version of the standard DEB animal model, with the specific aim to provide a tractable model with an explicit mass balance. Being a DEB-based model, DEBkiss follows an energy budget where resources from food are allocated to energy-requiring processes (growth, reproduction, maintenance, etc.). In contrast to DEBtox, the energy budget is explicit, which facilitates the incorporation of other traits (e.g., feeding and respiration, and embryonic development) and more extensive TK (e.g., dealing with losses due to reproduction). Chemical uptake is covered by a one-compartment TK model, accounting for growth (no measured body residues are needed). Chemical stress affects one (or few) of these processes in the energy budget, leading to specific patterns of effects over the life cycle.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Can be used at Tier 1.
Model purpose	Scientific / Regulatory	Originally intended for educational purposes (to introduce DEB theory in a simpler fashion), but turns out to be useful for many scientific questions. Not mentioned in regulatory context yet.
Questions / processes	Individual effects Effect propagation	Model aim is to explain growth and reproduction (and survival) of individuals over time, including the effect of toxicants. Can be extended to other endpoints such as feeding and respiration. Can also be linked to population-level calculations; could form a a simple building block for IBMs.
Environmental domain	Generic	No limitations in terms of taxa, chemicals or environmental compartment. Due to the absence of a 'reserve' compartment probably most useful for smaller invertebrates and fish.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Little-known	DEBkiss is a relatively new offspring from DEB theory, and therefore has received little attention. However, it is based on many of the well-established principles of DEB.
Public availability	Software extension	Implementations in Matlab and OpenModel.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction	The model is a simplification following from Dynamic Energy Budget theory, removing the 'reserve' compartment completely from the model.
Endpoints	Survival Reproduction	Statistics that can be calculated are LCx,t or ECx,t for any effect size x and time point t, or effects as a consequence of any exposure profile. Can be linked to population models.

Criteria	Categories	Comments
Space	No spatial context	
Time	Dates	Generally days – weeks – months.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario (both for calibration and predictions), as long as the exposure profile is known. Basic mixture effects can easily be added from related DEB-based models.
Abiotic environment	Food limitation Temperature	Food is integral part of all energy-budget models. Temperature affects all rate constants using the Arrhenius relationship.
Biotic environment	None	DEBkiss focusses on the individual level; biotic processes such as competition would need additional assumptions/modules.
Individuals	Stochastic Energy budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	-	DEBkiss deals with individuals, but can be linked to population approaches.
Calibration	Laboratory data	DEBkiss models are fitted to experimental data, and thus calibrated for each case.
Programming language	Matlab	Matlab and OpenModel implementations are available. For Matlab, within the BYOM framework: http://www.debtox.info/byom.html.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication Website	Concepts and equations are explained in a free e-book, basic manuals are available for the software implementations.

Assessment

Criteria	Description
Strengths	Can be calibrated using laboratory tests (most useful: partial life-cycle tests); simple model; integrates information on different endpoints over time, and extrapolate to other exposure scenarios.
Theoretical uncertainties	Based on a rigorous simplification of animal energetics over the entire life cycle. Inevitably, details on life history will be lost. The removal of 'reserve' from the model is a substantial deviation from DEB theory.

Criteria	Description
Empirical uncertainties	Many species, or data sets, require adaptations to the basic model; often difficult to find a unique mechanism of action of a chemical; no applications to birds and mammals.
Parametric uncertainties	Often difficult to identify all parameters from standard data sets.
Temporal uncertainties	Parameters assumed to remain constant over time.
Conclusions	DEBkiss can directly be used to estimate NEC or ECx values from datasets (including tests where exposure varies), and thereby improves upon the standard dose-response analyses. In extrapolation to untested situations, additional uncertainties need to be considered. In comparison to DEBtox, more details on the species are needed, but an explicit mass balance is used.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman and Metz (1984b)	Generic *Daphnia*		Predecessor of DEBkiss, sharing much of the same features, but lacking embryo stage.
Jager et al. (2013)	Pond snail	Food stress	General publication on the model, deriving the equations from a set of assumptions.
Jager (2018)	Pond snail	Food stress	Freely downloadable e-book, including, and expanding on, the original publication.
Barsi et al. (2014)	Pond snail	Acetone	Extension of DEBkiss to toxicant effects. Include effects on embryonic stages.
http://www.debtox.info/about_debkiss.html			General web page
Model applications			
Jager et al. (2014b)	Nematodes	Cadmium and fluoranthene	Extension to mixture effects. Include modification for initial slow growth in nematodes.
Jager et al. (2016b)	Sea urchin larvae	pH stress	Semi-quantitative use to tie together various endpoints.
Cedergreen et al. (2016)	Nematodes	Copper and temperature stress	Combined effects of chemical and temperature stress.
Jager and Ravagnan (2016)	Northern krill	Food stress	Parameterization and reconstruction of food history from field sampled animals.
Groeneveld et al. (2015)	Antarctic krill	Environmental conditions	DEBkiss used as module for krill life history in a population model.
Fiechter et al. (2015)	Chinook salmon	Environmental conditions	DEBkiss used as module for salmon life history in an ecosystem model.

Citations	Taxa	Chemicals	Comments
Rinke and Vijverberg (2005)	*Daphnia*	Environmental conditions	Model that is very similar to DEBkiss, but lacks embryonic phase and takes assimilation efficiency and maintenance rate as function of food density.
http://www.debtox.info/debkiss_appl.html			Full list of papers using DEBkiss models:

1.2.2.5 Net-Production Models (NPM, Brett and Groves 1979)

This modelling framework described in a textbook by Brett and Groves (1979) formulates an energy budget differently from DEB theory: Maintenance costs are deducted from the assimilated energy flux first, after which the remainder is allocated to growth and reproduction. The allocation between growth and reproduction is either fixed or variable (several empirical functions are used in different models). A wide range of models is used based on the same principle, but different in detailed formulation of the various processes and allocation rules.

General Properties

Criteria	Categories	Comments
Biological level	Individual	Generally intended as building block for population models.
Model purpose	Scientific / Regulatory	These models have a broad range of applications, both scientific and applied (e.g., in managing fish stocks). Usually as individual-level model in population calculations.
Questions / processes	Individual effects	These models provide an energy budget for the individual, often as part of a population model. NPMs describe how individuals use food to fuel growth and reproduction over their life cycle.
Environmental domain	Generic	These models are generic. They are commonly applied for invertebrates and fish.
Taxon specificity	Generic	
Toxicant specificity	Generic	Application of NPM models to toxic stress has been very limited.
Application	Established in science	Well-known in science, and used in managing fish stocks. Not used in chemical regulatory settings, as far as we know.
Public availability	-	Many implementations have been made. A standalone software with 33 fish models is available (not free): http://limnology.wisc.edu/research/bioenergetics/bioenergetics.html

Variables and Parameters

Criteria	Categories	Comments
Entities	Growth Reproduction	These models describe growth and reproduction of individuals as function of the food availability (and potentially other environmental factors).
Endpoints	Reproduction	Output of these models is growth and reproduction of individuals over time, under time-varying food conditions.
Space	No spatial context	NPM models do not have a spatial context themselves, but they can be used as building block in spatially explicit population models.
Time	Dates	

Criteria	Categories	Comments
Exposure / effects		These models are not usually applied for toxic stress. In cases where they are (e.g., Johnston et al.), the exposure concentration is directly linked to the stress level (i.e., no TK, although this could be included)
Abiotic environment	Food limitation	Feeding is explicitly modelled in NPMs; the energy from food is used to fuel growth and reproduction.
Biotic environment	None	NPMs deal with the life cycle of individuals. The population models in which they are implemented might include various interactions with other individuals.
Individuals	Energy budget	Individuals are described by an energy budget, different from that used in DEB theory. Maintenance costs are paid from the assimilated energy first, after which the remaining energy is allocated to growth and reproduction (and storage). The various models differ mainly in the treatment of the storage compartment and the rules for allocation between growth and reproduction. Generally, these models contain quite a number of descriptive elements (e.g., allometric functions).
Populations	-	These models have been implemented into various types of population models (i.e., DDE, PDE, IBM).
Calibration	Laboratory data	Generally, these models require rather detailed data to calibrate them to a specific species.
Programming language		Various

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Field data	General overview of validation status in Chipps and Wahl (2008). They conclude that agreement between predictions and data is generally poor, but the conceptual basis of the models is valid.
Sensitivity analysis	-	It is likely that sensitivity/uncertainty analysis has been performed on some NPM's. E.g., sensitivity analysis in Ananthasubramaniam et al. (2015).
Uncertainty analysis	-	
Documentation	Scientific publication	Many different models are available. The Sibly et al. (2013) model has been applied in an IBM (Johnston et al. 2014a), where also a detailed description in ODD/TRACE format is supplied. A software with 33 models is available (not free) with a manual.

Assessment

Criteria	Description
Strengths	Classical approach to bioenergetics in ecology, and therefore well established. They (should) explicitly deal with a mass balance for the individual, and follow how food is used to fuel growth and reproduction. Therefore, they are a useful tool for individual-based population models.
Theoretical uncertainties	Many different models exist, which are usually rather species-specific. There seems to be no attempts to provide an over-arching framework for all species. As these models generally contain a number of descriptive components, the extrapolation to other conditions is uncertain.
Empirical uncertainties	
Parametric uncertainties	In a simplification to an energy budget, many processes are lumped into energy/mass fluxes, which inherently implies uncertainties. In some cases, allocation functions are fitted to data.
Temporal uncertainties	
Conclusions	NPM models are widely applied in ecology, but have had very little application in ecotoxicology. Many different models using the net-production principle have been developed, making it difficult to select a single one as 'most useful'. Generally, these models rely on descriptive functions for various processes, which implies the need for substantial data sets for calibration, and raises questions on the ability to extrapolate beyond the tested conditions.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Sinko and Streifer (1969)	*Daphnia*	None	One of the earliest NPMs. Linked to a partial-differential equation to extrapolate to population dynamics.
Lika and Nisbet (2000)	Generic	None	A completely specified model, including the embryonic stage.
Brett and Groves (1979)	Fish	None	General book chapter explaining the bioenergetics approach.
Model applications			
Chipps and Wahl (2008)	Fish	None	Review of the validation status of fish bioenergetics models.
Nisbet et al. (2010)	*Daphnia*	None	Comparison with standard DEB theory. The NPM is included in a population context using delay-differential equations.

Citations	Taxa	Chemicals	Comments
Sibly et al. (2013)	Generic	None	Formulation of a specific NPM; allocation between growth and reproduction based on forced von Bertalanffy growth.
Johnston et al. (2014a)	Earthworms	Copper and chlorpyrifos	Inclusion of the Sibley et al NPM into an IBM and adding toxicant stress (this model has its own entry as population model).
Ananthasubramaniam et al. (2015)	*Daphnia*	Hypothetical	Dedicated *Daphnia model*, with details of moulting. Sensitivity analysis and some validation. Explicitly intended for ecotoxicology, but not fitted to toxicity test data.

1.2.3 Static Models

1.2.3.1 Stress Addition Model (SAM, Liess et al. 2016)

SAM (Liess et al. 2016b) predicts how the dose-response curve (for the endpoint survival) of a toxicant changes with the presence of an additional stressor that causes a given (fixed) mortality. SAM assumes that each individual has a general stress capacity towards all types of environmental stress, and that an organism dies if the general stress exceeds this capacity. The amount of stress capacity is considered to be normally distributed among individuals of a population. Experimentally observed mortality caused by a single stressor is converted into units of the hypothesized general stress. For this, a non-linear link is calibrated that relates the normally distributed general stress capacity (density function) to the non-normally distributed observed dose-response curve (cumulative distribution function). The general stress produced by the stressors when being applied alone is added, and is converted back to mortality. Therefore, in SAM the hypothesized general stress of different stressors is additive, but the mortality is not due to the non-linear link between general stress and mortality. Specific links have been established for various stressors and studies, and a generic calibration of the model has been performed using the mean link over all case studies.

General Properties

Criteria	Categories	Comments
Biological level	Individual	
Model purpose	Scientific	
Questions / processes	Additional stressors Mixture toxicity	SAM predicts interactive acute lethal effects of two stressors with different modes of action (pesticide + environmental stressor or pesticide + another toxicant with different mode of action).
Environmental domain	Generic	
Taxon specificity	Generic	The model was calibrated to freshwater macroinvertebrates.
Toxicant specificity	Generic	
Application	Little-known	
Public availability	Web application R package	https://www.systemecology.de/indicate/

Variables and Parameters

Criteria	Categories	Comments
Entities	Stress capacity	
Endpoints	Survival	
Space	No spatial context	
Time	Static model	
Exposure / effects	Toxicant mixtures	The model may be applied to mixtures of toxicants with different modes of action.

Criteria	Categories	Comments
Abiotic environment	-	The model considers interactive effects of one additional abiotic stressor. Otherwise it assumes environmental conditions as present when generating the input data (individual dose-response curve).
Biotic environment	None	The model considers interactive effects of one additional abiotic stressor. Otherwise it assumes environmental conditions as present when generating the input data (individual dose-response curve).
Individuals	Stress capacity	
Populations	-	The population is considered as a beta distribution of general stress capacity.
Calibration	Laboratory data Mesocosm data	Aquatic macroinvertebrates.
Programming language	-	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	Generic calibration of the model to the mean of a large number of independent case studies. The generic calibration fitted the data reasonably well and showed that an additional stressor can decrease the LC50 of a toxicant by a factor of 10, and LC10 by a factor of 100. No validation with additional data.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	SAM is an alternative to the traditional CA (Concentration Addition) and IA (Independent Action) models for the prediction of direct effects of toxicant mixtures, and in contrast to those was successfully applied to interactive effects of other environmental stressors.
Theoretical uncertainties	While a (non-linear) link between concentration and hypothesized general stress is established, no such link is established for the effect of the additional stressor (identity link assumed).
Empirical uncertainties	
Parametric uncertainties	The general stress capacity was hypothesized to be beta distributed within a population, but might follow a different distribution.
Temporal uncertainties	Static model, no prediction of effects dependent on varying exposure times (predictions only for the same exposure time used to generate the input data). Also no consideration of varying stress levels over time.

Criteria	Description
Conclusions	SAM is an empirical model to predict how sensitivity to direct effects increases through additional abiotic stress with different mode of action. The model cannot predict physiological interactions of stressors with the same mode of action. The mechanisms are hypothetical and not tested, but the predictions are in good accordance with observations.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Liess et al. (2016b)	Various aquatic vertebrates and invertebrates	Various pesticides	Original publication. Predictions of combined mortality from 1 pesticide + 1 additional environmental stressor.
Model applications			
Liess et al. (2020)	*Daphnia magna*	Esfenvalerate, prochloraz	Application to combined effects of two insecticides with different modes of action + food limitation stress.

1.3 Population Level Models

1.3.1 Discrete Models

1.3.1.1 Simple Discrete Models (Calow et al. 1997)

Discrete population models proceed in discrete time steps of e. g. 1 day, 1 year or 1 generation and use recurring difference equations to describe population size at the different modelled time steps. The example of Calow et al. (1997) uses a rewritten form of the discrete Euler-Lotka equation. The model uses a few key traits of a species: time and survival probability for juveniles to reach adulthood, and time and survival probability between breeding attempts, and the number of female offspring in each breeding attempt. The traits may be influenced by stressors (which have to have a constant effect on these traits over the life cycle). The traits are translated into a population multiplication rate (or growth rate). The authors demonstrate how the model is adapted (simplified) to deal with particular life histories (e.g., semelparous vs. iteroparous).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Regulatory	Objective of the authors is to increase the relevance of environmental risk assessment by introducing a simple population approach.
Questions / processes	Effect propagation	Effects on several key individual life-history traits are propagated to a population growth rate (or, in fact, a multiplication rate).
Environmental domain	Generic	
Taxon specificity	Generic	The model is not specific for any particular species.
Toxicant specificity	Generic	The model is not specific; it can be used for any chemical (or other stressor).
Application	Well known in science	The model has quite a few applications in scientific studies.
Public availability	-	There does not seem to be a publicly available version, but the model can be easily implemented in many software applications (incl. Excel)

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Reproduction	Mortality is included as survival probability in the juvenile stage and between breeding attempt. Reproduction is included as number of female offspring per breeding attempt. Growth rate is implicitly included in the time needed to reach adulthood.
Endpoints	Population size	Population multiplication or growth rate is the output of the model. This represents the long-term growth rate of the population in a constant environment.

Criteria	Categories	Comments
Space	No spatial context	
Time	Generation times	Time is included as the length of the juvenile period and the interval between breeding attempts.
Exposure / effects	Chronic vs. pulse	The model is limited to long-term constant chemical exposure.
Abiotic environment	-	No description of the environment. Life-history traits are necessarily constant.
Biotic environment	-	No biotic environment whatsoever; the model output is the population growth under constant conditions.
Individuals	Homogeneous	The population is divided into two classes: juveniles and adults. Within a stage, all individuals are identical.
Populations	Other	Population is structured in two stages.
Calibration	Laboratory data	In general, the model will be calibrated from laboratory data on the life-history traits and the effects of chemicals on these traits.
Programming language	-	Not mentioned.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	Yes	Plots are made to assess the sensitivity of the different parameters on the multiplication rate, and how this sensitivity differs between radically different life-history types.
Uncertainty analysis	No	
Documentation	Scientific publication	The model is extremely simple, and well documented.

Assessment

Criteria	Description
Strengths	Extremely simple extrapolation from individual traits to a relevant population-level statistic. Due to a lack of ecological complexities, the results are easy to interpret.
Theoretical uncertainties	The model relies on constant conditions, which implies that it is only relevant for long-term constant exposure, and only relevant when the stress level on the individuals is constant (i.e., no toxicokinetics). The model includes no environment (no food, no predators, etc.) and no intraspecific interaction. All individuals within a stage are taken as identical.
Empirical uncertainties	-
Parametric uncertainties	All toxicokinetics and -dynamics reduced to an immediate and constant effect on the vital rates.
Temporal uncertainties	The environment needs to remain constant (otherwise, there is no constant multiplication rate).

Criteria	Description
Conclusions	The model is extremely simple, and useful to provide a quick assessment of the population consequences of certain stressors. However, as PPP exposure is not constant, the usefulness within this context is limited.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Calow et al. (1997)	Generic	Generic	The main publication that was reviewed here.
Calow and Sibly (1990)	Generic	Generic	First publication of the rewritten Euler-Lotka equation (in the appendix).
		Model applications	
Hanson (2011)	Roach	Generic	Population response to effect on various traits; comparison of three models (unstructured, two-stage and age structured).
Hanson and Stark (2011)	*Daphnia*	Generic	Comparison to more complex age-structured matrix.
Ramskov and Forbes (2008)	*Capitella* (polychaete worm)	Sediment OC	
Pedersen et al. (2009)	Freshwater snail	Polycyclic musk	
Widarto et al. (2004)	Earthworm	Nonylphenol	
Salice and Miller (2003)	Freshwater snail	Cadmium	Extended to a three-stage model (embryo stage added), but also a very similar direct derivation of the multiplication rate from stage durations, survival probabilities and fecundity.

1.3.1.2 Probabilistic Discrete Models (Fabre et al. 2006)

Fabre et al. (2006) published a state-space model for integrated pest management strategies against the aphid *Rhopalosiphum padi*, the main vector of the *Barley yellow dwarf virus* (BYDV). The model requires a single early assessment of the proportion of plants infested by aphids as input. Then, the model predicts the percentage of plants infested by *R. padi* during autumn (a predictor of the need for insecticide sprays against BYDV vectors). Population development proceeds in discrete time steps of 1 day and is temperature-dependent, but density-independent (based on exponential growth). A Bayesian approach of statistical inference (state-space or integrated population modelling framework) is used to estimate uncertainty in the observation process as well as in the modelled growth process of aphids, so that the model predicts not a single value, but a whole probability distribution for the percentage of infected plants (i. e., model predictions are not deterministic but probabilistic).

General Properties

Criteria	Categories	Comments
Biological level	Population	Aphids (*Rhopalosiphum padi*)
Model purpose	Regulatory	Aim was to improve integrated pest management strategies against the aphid *R. padi*, the main Barley yellow dwarf virus (BYDV) vector in winter cereals during autumn in Europe.
Questions / processes	Others	The model is based on a temperature-dependent simulation of *R. padi* population dynamics. The model requires a single early assessment of the proportion of plants infested by aphids.
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	Aphids (*R. padi*) in winter cereals.
Toxicant specificity	-	It is not a toxicant, but a disease-vector model.
Application	Little-known	
Public availability	No	The statistical framework of the modelling approach is well known.

Variables and Parameters

Criteria	Categories	Comments
Entities	Growth	Exponential, temperature dependent growth.
Endpoints	Population size	Risk index of plant infestation.
Space	No spatial context	The field data were sampled on plots of 3 x 11 m; the state-space model is non-spatial.
Time	Days	Field data sampling was conducted from mid-September to end of November in 5-14 days intervals.
Exposure / effects	-	Growth of the aphid depending on initial conditions of infestation rate at t = 0.
Abiotic environment	Temperature	Growth rate was assumed to be temperature-dependent and to follow a linear day-degree model.
Biotic environment	-	

Criteria	Categories	Comments
Individuals	Homogeneous	
Populations	Exponential growth	
Calibration	Field data	Parameter estimation from field data (1989-1994).
Programming language	Turbo Pascal with Borland Delphi 6.0; WinBUGs	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Validation with independent field data from 1995-1999 at the same sites.
Sensitivity analysis	Yes	Inherently done in Bayesian approach as the posterior distribution for each estimated parameter is given.
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	This is a stochastic model yielding risk of plant infestation by aphids based on field data.
Theoretical uncertainties	Simple exponential growth depending on temperature assumed. No inter- or intraspecific processes included. Authors conclude themselves that the moderate coefficient of determination suggested that many other factors drive abundance of aphid populations in the field.
Empirical uncertainties	Temperature and initial infestation can be measured; however, sampling needs to be done thoroughly to avoid mis-counting and decision-makers may undoubtedly require some basic training and instructions on making such observations.
Parametric uncertainties	High: simple exponential growth model depending on temperature and initial conditions as only parameters.
Temporal uncertainties	Only a short time period has been observed, in which the system is assumed to be stationary.
Conclusions	Since the model is parameterized, risk could be extrapolated to other.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Fabre et al. (2006)	aphids (*Rhopalosiphum padi*)	None; disease (infestation) risk	

1.3.1.3 Discrete Models with Multiple Stages (Gledhill and van Kirk 2011)

Gledhill and Van Kirk (2011) present a generic approach for the simulation of long-term freshwater fish population dynamics. With appropriate parameterization, this model is applicable for many fish species. The model runs in discrete yearly time steps and simulates a population that is built up by cohorts of a couple of age classes with different characteristics, leading over to the special case of matrix models among discrete models (see section 1.3.1.4). It provides information on how the size of a contaminated population behaves (increase, constant, decrease) as compared to an unaffected control population. As endpoint the model focuses on equilibrium population size instead of population growth rate which is often used in other approaches. The cohorts (age classes) are characterised by equations considering the life cycle traits mortality, population density, reproduction and growth. For the modelling of lethal toxic effects, an LC50 can be incorporated in a log-logistic concentration-response curve.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	Risk assessment of chemicals in freshwater fish species.
Questions / processes	Effect propagation	Extrapolation from individual-level to population-level toxicity effects and describing the influence of potential variability in natural conditions.
Environmental domain	Freshwater	
Taxon specificity	Generic	The model is generic for many freshwater fish species. Application to bluegill sunfish (*Lepomis macrochirus*).
Toxicant specificity	Generic	Model application to selenium in bluegill sunfish.
Application	Little known	The model has quite a few applications in scientific studies.
Public availability	-	No information. Equations shown in publication.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Population size Reproduction	Mortality is described by the annual survival rate and additionally increases with high population density. The model includes aging of individuals and describes altering individual characteristics (e.g. sensitivity to density-dependence), depending on the current age of an individual. Individuals may reproduce if they reach the age of two years. Growth is described with a von Bertalanffy growth model and fecundity depends on body length.
Endpoints	Population size Survival	Population size (equilibrium or extinction) has been considered the most important and the intrinsic population growth rate (PGR) has been considered as the second relevant model output.
Space	No spatial context	
Time	Years	The model proceeds in time steps of 1 year.

Criteria	Categories	Comments
Exposure / effects	Varying concentrations	Toxicity is modelled with a given total body toxin concentration that increases survival of larvae and juveniles using a log-logistic dose-response relationship. Ingestion of toxicants or chemical fate is not explicitly considered.
Abiotic environment	Food limitation	Competition for resources is indirectly included by density-dependent effects.
Biotic environment	Intraspec. competition	Intraspecific competition for resources is indirectly included by considering density dependence. Mortality caused by fishing is implicitly included in the parameterization of background mortality.
Individuals	Stochastic	Stochasticity is incorporated in all survival and growth rates to consider natural variability. The population is modelled as an amount of cohorts, not of individuals.
Populations	Logistic growth	No absolute population size, but relative population size is described in relation to the environmentally determined population size (possible population size without consideration of toxic effects).
Calibration	Laboratory data Field data	Parameterization for *Lepomis macrochirus* was implemented with published results found in literature.
Programming language	-	Not mentioned.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Field data	Model outputs compared to field observation data and laboratory test results. Conclusions are difficult because of different requirements and lacking of data.
Sensitivity analysis	Yes	S_b (density-independent first year survival rate) was fixed at several values to investigate model sensitivity to Sb. The model reacts sensitive to changes in S_b and is assumed to be relatively insensitive to changes in R (reproductive potential).
Uncertainty analysis	No	
Documentation	Scientific publication	Well structured publication, but the large amount of mathematical equations makes it hard to understand.

Assessment

Criteria	Description
Strengths	Laboratory data (e. g. LC50) can be used for parameterization. The model is generic and can be applied to any freshwater fish species. Consideration of life-cycle characteristics and example for the incorporation of toxicology.
Theoretical uncertainties	No abiotic conditions and interspecific relationships explicitly considered. No explicit incorporation of chemical fate or toxicokinetics and toxicodynamics.

Criteria	Description
Empirical uncertainties	R_{Sb} (population recruitment potential) and S_b (density-independent first year survival rate) are difficult to observe.
Parametric uncertainties	Not considering individuals but cohorts aggregates across various processes at the individual level which increases uncertainty.
Temporal uncertainties	Assumption of the environment to be closed.
Conclusions	A model is a highly simplifying approach that is potentially applicable to various fish species. The long time steps of 1 year and the focus on long-term effects (changes in equilibrium population size) may be useful for the assessment of chronic effects, but render the model unsuitable for a detailed assessment of short-term effects.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Calow et al. (1997)	Generic	Generic	The main publication that was reviewed here.
Calow and Sibly (1990)	Generic	Generic	First publication of the rewritten Euler-Lotka equation (in the appendix).
		Model applications	
Hanson (2011)	Roach	Generic	Population response to effect on various traits; comparison of three models (unstructured, two-stage and age structured).
Hanson and Stark (2011)	*Daphnia*	Generic	Comparison to more complex age-structured matrix.
Ramskov and Forbes (2008)	*Capitella* (polychaete worm)	Sediment OC	
Pedersen et al. (2009)	Freshwater snail	Polycyclic musk	
Widarto et al. (2004)	Earthworm	Nonylphenol	
Salice and Miller (2003)	Freshwater snail	Cadmium	Extended to a three-stage model (embryo stage added), but also a very similar direct derivation of the multiplication rate from stage durations, survival probabilities and fecundity.

1.3.1.4 **Simple Matrix Models (Charles et al. 2004)**

Matrix models are a special case of discrete models that use matrix algebra for demographic population studies in which populations are classified in multiple life stages or age classes (Soetaert and Herman 2009). The reviewed model of Charles et al. (2004) is a typical representative of the Leslie matrix models. A population is represented by a number of subpopulations that represent different life stages (age classes or developmental stages). Each life stage has a subpopulation size, a probability for survival of a time step, and a probability for the transition to another life stage during a time step. The survival and transition probabilities can be constants or functions of the environment. Only females are modelled as males do not produce offspring. When fertility rates are low, the population growth rate (Lambda) and the distribution of subpopulation sizes among the life stages converges to a stable equilibrium after a number of time steps with fixed transition probabilities. When fertility rates are high, the life stage distribution oscillates with a period of 1 generation time.

In this example, a matrix with life stages and age classes was calibrated to Chironomidae populations in the laboratory. Every time step the surviving individuals moved from one age class to the next. The duration of the egg stage and the first three larval stages (L1, L2, L3, in number of age classes = time steps) was fixed. Duration of L4 (including the pupal stage) and the fecundity of adults were functions of the amount of provided food. Survival probabilities of the larval stages were fixed. Without food limitation, the model predicts a strongly oscillating population growth rate. The predicted effect of food limitation on the population growth was consistent with laboratory studies. The model correctly predicted the effect of food amount on the growth rate and generation time, and predicted a relative decrease of eggs and a relative increase of L4 in the life stage distribution with food limitation.

General Properties

Criteria	Categories	Comments
Biological level	Population	Matrix models are population models that consider demographic effects.
Model purpose	Scientific	Matrix models are also used in a regulatory context.
Questions / processes	Others	Demonstrating the general potential of matrix models for risk assessment, by showing that matrix models can predict the effects of environmental stressors on the population dynamics of the ecologically relevant group of Chironomidae.
Environmental domain	Freshwater	Matrix models can be applied to any group of organisms. The specific model of Charles et al. (2004) considers a freshwater species.
Taxon specificity	Taxon-specific	Matrix structure and transition probabilities must be adapted to the life cycle of a specific species.
Toxicant specificity	Generic	No toxicant included yet. Other matrix models include toxicant effects as well.
Application	Little-known	Matrix models represent a huge family of models designed for input data from life table response experiments (LTRE). The specific model from Charles et al. Is little-known and has been only published in a journal article.
Public availability	-	Formulas provided in a journal article. The book of Caswell (2001) is the standard work on these models.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality Reproduction	The model considers the size of subpopulations representing different life stages, the probability of survival in a life stage during a time step, and the probability to move from one stage to another during a time step.
Endpoints	Population size Population structure Survival Reproduction	Prediction of the population size, life stage distribution, probability of population extinction and the fecundity is possible.
Space	No spatial context	
Time	Days	
Exposure / effects	Chronic vs. pulse Varying concentrations Repeated exposure	No toxicant included in this example, but other publications deal with toxicant stress. Due to the subdivision in discrete classes, and the limited number of state variables for the individual (generally one: age or stage), these models are best suited for constant exposure.
Abiotic environment	Food limitation	Food limitation increases the duration of L4 and decreases fecundity of adults. The functions for the effect of food were obtained from non-linear regressions of food vs. duration (or fecundity) using data from a life table response experiment (LTRE).
Biotic environment	-	No species interactions considered. Other matrix models include forms of density dependence.
Individuals	Homogeneous	The TD model embodies a very simple energy budget: growth is the net result of biomass formation and losses due to respiration. Biomass formation is treated simply as a rate, which is influenced by various environmental factors. The organisms do not have a reserve, which implies that population growth responds immediately to changes in the environment.
Populations	Life stages	Each age class has a transition probability to move to the next class.
Calibration	Laboratory data	LTRE with female midges (Chironomidae) in laboratory cultures. Daily reproduction probability estimated as encounter probabilities between surviving male and unmated female adults.
Programming language	-	Not reported for this example, but many different matrix-model implementations exist.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	Without food limitation, the model predicts a high oscillating population growth rate. The predicted effect of food limitation on the population growth was consistent with laboratory studies. The model correctly predicted the effect of food amount on the growth rate and generation time, and predicted a relative decrease of eggs and a relative increase of L4 in the life stage distribution with food limitation. In laboratory cultures, population equilibrium is reached at low food levels mainly because of reduced of egg production and prolongated development time.
Sensitivity analysis	Yes	Elasticity analysis was performed without food limitation, but not informative due to the cyclicity of the matrix instead of a stable stage distribution. The effect of food limitation on the population growth rate depended particularly on the duration of L4 together with the pupal stage.
Uncertainty analysis	No	
Documentation	-	Structured journal article, but no explicit documentation for users.

Assessment

Criteria	Description
Strengths	Matrix models have been used for a long time and a number of efficient tools for the sensitivity analysis have been developed. Matrix models can be easily adapted to the specific life history of a species and predict demographic effects with low computing capacity.
Theoretical uncertainties	In matrix models, all biological processes are purely represented by transition and survival probabilities. These probabilities integrate numerous biological processes, each of which can be affected by the environment in an unpredictable way. The individuals are grouped in discrete classes, and generally only have one state variable (age or stage); a realistic representation of individuals in a changing environment often requires more state variables.
Empirical uncertainties	Transition and survival probabilities can be modelled as functions of environmental parameters. When using a simple function for each environmental parameter (as in Charles et al. 2004), the model considers only additive effects of the environment. However, several parameters may have interacting effects on such a probability. The correct calibration of such interactions is essential for the model output but requires an enormous amount of experimental data. Matrix models are therefore more suited for the analysis of experimental data under controlled conditions than for the prediction of population dynamics under highly variable field conditions.
Parametric uncertainties	Generic matrix models do not consider interactions of the modelled population with the biotic and abiotic environment. Stochastic processes and Allee effects that can drive small populations to extinction irrespective of the mean survival probabilities are typically not considered (as in Charles et al. 2004) but may be introduced using probability functions for the transition probabilities. Spatial heterogeneity cannot be considered.

Criteria	Description
Temporal uncertainties	Transition probabilities in matrix models can be a function of environmental parameters that can be changed during the simulation.
Conclusions	Matrix models describe population dynamics more precisely, but need more input data compared to growth models based on differential equations. Matrix models describe population dynamics less precisely than IBMs, but need less computing capacity. This makes them potentially useful as sub-models in complex ecosystem models, if enough data are available to calibrate the effects of the environment on the population dynamics. When computing capacity is not the limiting factor, IBMs can be more realistic with a similar amount of input data required.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Charles et al. (2004)	Chironomidae	None	Used a a case example for matrix models.
Caswell (2001)	Generic	None	Standard work on matrix modelling.
		Model applications	
Chandler et al. (2004)	Copepods	Fipronil	Stage-structured matrix model.
Spromberg and Meador (2006)	Fish: coho salmon (*Oncorhynchus kisutch*), sockeye salmon (*Oncorhynchus nerka*) and chinook salmon (*Oncorhynchus tshawytscha*)	PAH and PCB	Age-structured matrix model.
Hanson and Stark (2011)	*Daphnia*	Hypothetical	Leslie matrix with two types of density dependence; comparison to the simpler two-stage approach.
Hamda et al. (2014)	Springtails (*Folsomia candida*)	Cadmium	Stage-structured matrix model, including effects of a toxicant on vital rates, density dependence, and environmental stochasticity (on temperature).

1.3.1.5 Matrix Modelling with DEB (Klanjscek et al. 2006)

Use of the dynamic energy budget theory in individual-level TKTD models has been described in section 1.2.2.2. DEB describes survival, growth and reproduction over the life cycle (and the effects of stressors on these traits) and can be used as module for individual-level effects in population models. In the example of Klanjscek et al. (2006), a DEB model for animals is integrated into a matrix modelling context (see section 1.3.1.4). An IBM implementation was used for comparison.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	
Questions / processes	Effect propagation	Population dynamics (no toxicants included).
Environmental domain	Generic	The paper is a general presentation of the integration of DEB models into a matrix population model.
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Little-known	Both DEB and matrix modelling are well-established in science. Their combination is not too common (more work is done with kmDEB and DEBtox).
Public availability	None	No publicly available version.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Life history of individuals is dealt with by the DEB component, and propagation to population dynamics in the matrix model.
Endpoints	Population size	
Space	No spatial context	
Time	Dates	
Exposure / effects	-	No toxicant stress included.
Abiotic environment	Food limitation Temperature	Fluctuating food was simulated; temperature can easily be added.
Biotic environment	-	No interactions between individuals or with their food.
Individuals	Stochastic Energy-budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	Lfe stages	Matrix model, stage structured.
Calibration	-	No calibration, only simulations for a hypothetical animal.

Criteria	Categories	Comments
Programming language	Unclear	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation attempts.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	DEB is a well-established theory for individuals, and matrix modelling is popular in eco(toxico)logy.
Theoretical uncertainties	Apart from the uncertainties in DEB, the matrix model follows a population of one species in isolation (no interspecies competition, no predators, no parasites, no interaction with food etc.). Matrix models work with discrete classes or individuals, which means that individuals receive average properties. Furthermore, these models allow for only a single state variable of the animal (i.e., its stage), which hampers extension to more realistic settings.
Empirical uncertainties	Many species, or data sets, require adaptations to the DEB model; often difficult to find a unique mechanism of action of a chemical.
Parametric uncertainties	Often difficult to identify all DEB parameters uniquely from standard data sets.
Temporal uncertainties	
Conclusions	Matrix models work in discrete time and with classes of individuals (stages). This does not lend itself naturally to a combination with DEB, and does not allow for more realistic TKTD of individuals. The focus on a single population in isolation limits its realism for field situations.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Klanjscek et al. (2006)	Generic	Generic	General presentation of the DEB-matrix model.

1.3.1.6 Matrix Modelling with DEBtox (Billoir et al. 2007)

DEBtox has been described in section 1.2.2.3. As DEBtox describes survival and reproduction over the life cycle (and the effects of stressors on these traits), the output can be integrated in population models. In the examples of Lopes et al. (2005) and Billoir et al. (2007), DEBtox is integrated into a matrix modelling context (see section 1.3.1.4).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation	
Environmental domain	Generic	
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Little-known	Both DEBtox and matrix modelling are well-established in science. Their combination is not too common, although there are a number of applications (also with kmDEB and matrix models).
Public availability	-	No publicly available version.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Life history of individuals is dealt with by the DEBtox component, and propagation to population dynamics in the matrix model.
Endpoints	Population size	Applications focus on the population growth rate (i.e., the intrinsic rate of increase).
Space	No spatial context	
Time	Dates	
Exposure / effects	-	Constant exposure (the vital rates in the matrix are taken constant).
Abiotic environment	Food limitation Temperature	Constant food limitation and constant temperatures can easily be included, as in other DEB-based models.
Biotic environment	-	No interactions between individuals or with their food.
Individuals	Stochastic Energy-budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	Exponential growth	Matrix model, age structured (Leslie matrix).
Calibration	Laboratory data	The DEBtox model was fitted to experiments with *Chironomus* and *Daphnia* at the individual level.

Criteria	Categories	Comments
Programming language	Unclear	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation attempts.
Sensitivity analysis	Yes	Sensitivity of the population growth rate as a function of the toxicant, split up over the different vital rates.
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	DEBtox is a well-established model for analysing toxic effects on individuals, and matrix modelling is popular in eco(toxico)logy.
Theoretical uncertainties	Apart from the uncertainties in DEB, the matrix model follows a population of one species in isolation (no interspecies competition, no predators, no parasites, no interaction with food etc.). Matrix models work with discrete classes or individuals, which means that individuals receive average properties. Furthermore, these models allow for only a single state variable of the animal (i.e., its age), which hampers extension to more realistic settings.
Empirical uncertainties	Many species, or data sets, require adaptations to the DEBtox model; often difficult to find a unique mechanism of action of a chemical; no applications to birds and mammals.
Parametric uncertainties	Often difficult to identify all DEB parameters uniquely from standard data sets.
Temporal uncertainties	
Conclusions	Matrix models work in discrete time and with classes of individuals (age classes). This does not lend itself naturally to a combination with DEB, and does not allow for more realistic TKTD of individuals. The focus on a single population in isolation limits its realism for field situations.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Lopes et al. (2005)	*Chironomus*	Methiocarb	Only the hazard model (endpoint survival) from DEBtox was used.
Billoir et al. (2007)	*Daphnia*	Cadmium	Complete DEBtox model used.
		Model applications	
Billoir et al. (2009)	*Moina* (Cladocera)	Toxic cyanobacteria	
Ducrot et al. (2007)	*Valvata* (snail)	Zinc	Matrix model with two stages (juveniles/adults).
Smit et al. (2006)	Amphipods	Contaminated sediments	Matrix based on size classes, some validation to field data.

1.3.1.7 Matrix Modelling with kmDEB (Klok and De Roos 1996)

Kooijman-Metz (km) DEB is a predecessor of DEB theory, and bares a close resemblance to DEBtox. It is based on a simplified energy budget. The output in terms of growth, survival and reproduction is linked to a simple matrix model. In comparison to DEBtox, kmDEB has no reserve, no maturity maintenance, and no TK considerations. Klok and De Roos (1996) provide an example of how a kMDEB module can be integrated in a matrix population model (see section 1.3.1.4).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Individual effects Effect propagation	Translation of effects in individuals to population dynamics, both under constant and pulsed exposure.
Environmental domain	Generic	The model is generic although only earthworms are discussed in the papers of Klok et al (see below).
Taxon specificity	Generic	
Toxicant specificity	Generic	Hypothetical compounds are used; no toxicokinetics, so instant steady state.
Application	Little-known	kmDEB is a predecessor of DEB theory, and closely resembles DEBtox. Here, it is linked to a matrix model to calculate population growth rate.
Public availability	-	Unclear; the model does not seem to be publicly available.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Effects on individuals are treated in the kmDEB module, and they are propagated to a population growth rate using the matrix model.
Endpoints	Reproduction Population size	Endpoint is the exponential growth rate of the population, and stage distribution, under constant conditions.
Space	No spatial context	
Time	Dates	The matrix calculation requires growth, survival and reproduction over one life cycle, and translates this into a population growth rate.
Exposure / effects	Repeated exposure Chronic vs. pulse exposure Varying concentrations	Only chronic, constant, exposure. This limitation is caused by the focus on the population growth rate, and partly by the kmDEB module which lacks a TK module.

Criteria	Categories	Comments
Abiotic environment	Food limitation	Different food levels can be used in the calculation, as long as they remain constant over time.
Biotic environment	-	No biotic factors, apart from the individual's energy budget.
Individuals	Energy budget	Energy budget for growth and reproduction.
Populations	Exponential growth	Exponential growth rate is calculated.
Calibration	Laboratory data	Model is fitted to data; the population predictions do not require additional parameters.
Programming language	Unknown	There is no up-to-date implementation of this model, though it would be easy to program in many software applications.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	Some general comparisons to field data. Model is based on first principles and describes life history patterns quite well.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	Matrix models are well described in many places.

Assessment

Criteria	Description
Strengths	The matrix calculation is simple and easy to interpret due to the limited ecological complexity. The intrinsic rate of increase is useful as a fitness measure for the population under stress.
Theoretical uncertainties	The kmDEB model is not used anymore, as it has been replaced by DEBtox. The focus on population growth rate assumes a constant environment over many generations, which is unrealistic.
Empirical uncertainties	
Parametric uncertainties	The original matrix model was stage-structured, which implies that individual growth is represented as a probability to move to the next stage.
Temporal uncertainties	The population growth rate calculation is based on a constant environment over many generations. This is unrealistic.
Conclusions	Matrix models are a simple way to extend the DEB-based TKTD model to the population level. The kmDEB model is outdated and has been replaced by DEBtox.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Klok and De Roos (1996)	Earthworms	Copper	
		Model applications	
Klok et al. (1997)	Earthworms	Copper	Basically the same as Klok and De Roos (1996)
(Klok et al. 2006)	Earthworms	Heavy metal mixture	Model used to translate bioassay from field-collected soil to population level.
Klok et al. (2007)	Earthworms	Copper	
Klok (2008)	Earthworms	Zinc	Interaction of zinc and population density.
Jager and Klok (2010)	Earthworms	Copper	Comparison of different DEB models and simple population approaches.

1.3.2 Continuous Models

1.3.2.1 Simple Ordinary Differential Equations (ODE) Models (Barnthouse 2004)

In contrast to discrete difference equation models, ordinary differential equation (ODE) models deal with time as a continuous variable (Soetaert and Herman 2009). If the differential equations cannot be solved analytically (which is the case for most of the realistic model applications), they are solved numerically by proceeding in small time steps, which reminds on the way discrete models proceed. However, ODE models are solved in very small time steps which do not need to be constant but can be adapted to the local stiffness of an equation (the amount of change in a state variable with change in time). Due to their simplicity, generic ODE models are widely used in ecology to describe population growth and recovery after disturbance. They are often linked to simple dose-response models for individual-level effects in ecotoxicology.

Here we reviewed a well-known example of Barnthouse (2004). The model views agrochemical application as a disturbance that eliminates a certain fraction of the individuals in a population. Instantaneous growth rates (from life table data, mesocosm data, or guessed from generation times) are used in conjunction with logistic growth to estimate the time needed to reach recovery (a certain percentage of maximum population size). The population is assumed to immediately start growing after the disturbance at maximum rate. Multiple disturbances can be simulated, and the effect of immigration was included.

General Properties

Criteria	Categories	Comments
Biological level	Population	Could be used at Tier 1, due to simplicity and low data demands.
Model purpose	Scientific / Regulatory	The model is intended to discuss/illustrate the issue of recovery in the risk assessment of PPPs.
Questions / processes	Population recovery	Aim of the model is to predict recovery rates based on very simplistic assumptions regarding the species' life history and chemical effects.
Environmental domain	Freshwater	The model is applied to freshwater organisms (algae, invertebrates, vertebrates). In principle, application to the terrestrial environment is possible, although the assumption of an immediate start of recovery after pesticide application would need to be evaluated.
Taxon specificity	Generic	
Toxicant specificity	Generic	Assumption is made that recovery starts at maximum rate directly after the pesticide application, so this model does not apply to chemicals that are (even slightly) persistent.
Application	Little-known	The model does not seem to have been applied a lot, but has a good citation score, so has managed to stir some discussion on the topic of recovery. WoS: 52 citations 24/6/2016.
Public availability	-	There does not seem to be a publicly available version, although the model would be easy to implement (e.g., in Excel).

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality	
Endpoints	Recovery time	The model estimates recovery time of a population to a certain percentage of maximum population size, starting from a disturbance that reduces population size to a certain fraction of the maximum.
Space	No spatial context	
Time	Dates	
Exposure / effects	Repeated exposure	Exposure is treated as an instantaneous disturbance, reducing population size. Multiple disturbances can be included.
Abiotic environment	None	No description of the environment.
Biotic environment	None	No explicit biotic interactions. The population grows according to a logistic function.
Individuals	Homogeneous	This is an unstructured population approach: the population grows according to the logistic function.
Populations	Logistic growth	
Calibration	Laboratory data Mesocosm data Field data	Range of freshwater species was used. Various sources of data can be used to estimate the intrinsic rate of population increase that is needed for the logistic growth function.
Programming language		Not mentioned, but can easily be implemented in various software (incl. Excel).

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Mesocosm data	Several groups of invertebrates were used. Very crude comparison between recovery rates observed in mesocosms and those predicted from the model. Due to the high variability in the observations, no general conclusions can be drawn.
Sensitivity analysis	No	
Uncertainty analysis	Yes	Very limited: The variation in the estimated values for the intrinsic rate are propagated to variation in recovery times for a few groups.
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	The model is very simple, very transparent, and easy to calculate. Nevertheless, it gives some insights into the vulnerability of various taxonomic groups to population reductions due to chemical stress. It focusses not on the toxic effects of the chemical but rather on the recovery of the populations.
Theoretical uncertainties	The model assumes that chemical stress only affects mortality, and that all effects are instantaneously reversed after the disturbance. No effects of other species (competitors, predators, etc.) on the recovery rates. Logistic growth is imposed. Growth rates are taken from data under optimal conditions, and therefore are best-case estimates.
Empirical uncertainties	It is difficult to estimate intrinsic rate of increase under realistic conditions.
Parametric uncertainties	Reducing population dynamics to intrinsic rate and logistic growth is rather crude.
Temporal uncertainties	The growth function remains the same over time.
Conclusions	A very simple model that illustrates the need to consider (or even focus on) recovery in the risk assessment of PPPs. The model lacks ecological realism, but can be used to calculate crude estimates for recovery time. Interestingly, the only chemical-specific input would be the reduction of the initial population size and the number and timing of the applications.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Barnthouse (2004)	Various freshwater organisms	General	

1.3.2.2 ODE Model Coupled to Species Size (Hendriks and Enserink 1996)

In this example of Hendriks and Enserink (1996), the basic parameters of an ODE mode on reproduction and survival were correlated to species size in order to make the model applicable to a wide variety of species. The model assumes exponential or logistic growth of the population. Toxicity is included by taking a standard log-logistic dose response and applying it to survival and reproduction rates (assuming a constant stress). In this way, effects can be expressed as percentage reduction in population growth rate of carrying capacity.

General Properties

Criteria	Categories	Comments
Biological level	Population	Could be used at Tier 1, due to simplicity and low data demands.
Model purpose	Regulatory	The model is intended as an improvement over classical approaches in environmental management.
Questions / processes	Effect propagation	Aim of the model is to predict impacts on population growth rate and carrying capacity, using dose response curves for survival and reproduction.
Environmental domain	Generic	Can be applied to all taxa.
Taxon specificity	Generic	
Toxicant specificity	Generic	Can be applied to all toxicants.
Application	Little-known	
Public availability	-	There does not seem to be a publicly available version, although the model would be easy to implement (e.g., in Excel).

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality Reproduction	
Endpoints	Population size	The model predicts effects on intrinsic rate of population increase or carrying capacity.
Space	No spatial context	
Time	Dates	
Exposure / effects	Constant exposure Dose-response	Exposure needs to be constant, otherwise effects on intrinsic rate and carrying capacity become meaningless.
Abiotic environment		No description of the environment.
Biotic environment	Intraspecific competition	No explicit biotic interactions. The population grows according an exponential or logistic function.
Individuals	Homogeneous	This is an unstructured population approach: the population grows according to the exponential or logistic function.
Populations	Logistic growth	

Criteria	Categories	Comments
	Exponential growth	
Calibration	Laboratory data Field data	Various sources of data can be used to estimate the parameters that are needed (including regressions based on body size). Was parameterized for *Daphnia* and cormorants.
Programming language	-	Not mentioned, but can easily be implemented in various software (incl. Excel).

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	The model is very simple, very transparent, and easy to calculate. Nevertheless, it gives some insights into the consequences of a reduction in survival and reproduction for the population level.
Theoretical uncertainties	The model assumes that the stress level on the organisms is constant throughout its lifetime. Therefore, the model is of little use for PPPs with highly time-varying exposure. The model imposes exponential or logistic growth, and does not include interactions with the environment (e.g, food and temperature) or with other species (e.g., predators and prey).
Empirical uncertainties	It is difficult to estimate intrinsic rate of increase under realistic conditions.
Parametric uncertainties	Reducing population dynamics to intrinsic rate and logistic growth is rather crude.
Temporal uncertainties	The model assumes a constant environment, including constant long-term toxicant stress.
Conclusions	A very simple model that illustrates how simple population approaches can aid environmental risk assessment. The model lacks ecological realism, but can be used to calculate consequences of effects on survival and reproduction in a transparent manner. The model is very similar to the approach taken by Barnthouse (2004).

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Hendriks and Enserink (1996)	*Daphnia* and *cormorants*	Various	

1.3.2.3 ODE Model for Aphids (Adams et al. 2005)

Adams et al. (2005) provide an example for a more complex ODE based model specifically designed for aphids. The model includes seven ODE's of varying complexity considering birth and death rate (potentially time- and/or state-dependent), exponential growth and death due to pesticides. Simultaneous exponential birth and death, density-dependent birth and death rates are incorporated in this standard logistic model. The focus of the paper is on finding the optimal statistical fit (parameter estimation) of the field data to the ODE models.

General Properties

Criteria	Categories	Comments
Biological level	Population	aphids
Model purpose	Scientific	Purely academic; To offer sophisticated mathematical observations based on a field study by Banks and Stark (2004) that explored the combined effect of vegetation diversity and chemical intervention; to understand the influence of natural enemies or other margin-based factors separately from and in interaction with that of the insecticide.
Questions / processes	Others	Fitting population dynamics models to data from the Banks–Stark field study using ODE's (ordinary differential equation) and fitting piecewise constant and piecewise linear time-varying coefficients in the corresponding non-autonomous ODEs.
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	Aphids (*Mycus persicae, Brevicoryne brassicae*)
Toxicant specificity	Toxicant-specific	Imidacloprid
Application	Little-known	
Public availability	-	ODE's are specified and solved with MATLAB ODE solver ode45 incl. Runge-Kutta solver

Variables and Parameters

Criteria	Categories	Comments
Entities	Population size	
Endpoints	Population size	Expressed as density: mean number of aphids / m3
Space	No spatial context	
Time	Days	
Exposure / effects	Repeated exposure	Three levels of insecticide spray (no, light, or heavy spray) applied once to broccoli patches surrounded by different margin types (bare or weedy ground). Imidacloprid spray was applied on July 23, August 13, and August 27, denoted by days 0, 21, and 35, respectively, in the paper by Banks and Stark (2004).
Abiotic environment	None	Parameterized with data from controlled field experiment.

Criteria	Categories	Comments
Biotic environment	None	Parameterized with data from controlled field experiment.
Individuals	Homogeneous	
Populations	Exponential growth Logistic growth	
Calibration	Field data	The whole paper is about calibration or parameter estimation.
Programming language		MATLAB

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	No	Not in terms of varying input parameters, but in respect to assessing the variance in the estimated model parameters by employing sensitivity equations to compute standard errors. Incorporation of time-varying coefficients in the models often yields a statistically significant improvement in fit.
Uncertainty analysis	No	See above.
Documentation	Scientific publication	Detailed description in the first publication following the ODD protocol. Comprehensive help functions on the website.

Assessment

Criteria	Description
Strengths	Rigorous statistical fitting procedure of field data to mathematical models.
Theoretical uncertainties	High: No processes on the individual level, no external stressors etc.
Empirical uncertainties	High: Models were fitted to data of a limited study.
Parametric uncertainties	Low: Extensive statistical fitting procedures applied.
Temporal uncertainties	High: Models were fitted to data of a limited study.
Conclusions	Cannot be used for risk assessment, as the estimated parameters only fit the specific field study.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Adams et al. (2005)	Aphids (*Mycus persicae*, *Brevicoryne brassicae*)	Imidacloprid	

1.3.2.4 **Staged ODE Model for Mosquitofish (Cabral et al. 2001)**

The population model of Cabral et al. (2001) was developed to assess the risk of the non-ionic surfactant Genapol OXD-080, a fatty alcohol polyglycol ether, to the non-target mosquitofish *Gambusia holbrooki* after application against crayfish (*Procambarus clarkii*) in rice paddies. In contrast to the previous examples, the population is modelled as consisting of three stages: immatures, females and males. Straightforward ordinary differential equations (ODEs) are used to model the population. Therefore, there seems to be a continuous flow of immatures to the mature stages. Intraspecific competition for food is included, and food availability and photoperiod vary over the year (as a table input).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Regulatory	Authors intended this paper as a contribution to the risk assessment for surfactants (to control damage by crayfish) in rice fields.
Questions / processes	Effect propagation	The model calculates population dynamics for mosquitofish when exposed to a specific surfactant.
Environmental domain	Freshwater	The model is parameterised for mosquitofish in rice fields, although model structure could be used for other purposes.
Taxon specificity	Taxon-specific	Mosquitofish
Toxicant specificity	Toxicant-specific	The model includes specific assumptions for the action of the surfactant Genapol OXD-080.
Application	Little-known	This specific model does not seem to have been applied apart from this single paper.
Public availability	-	There does not seem to be a publicly available version.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality Reproduction	The model considers maturation of immatures to males/females, reproduction, mortality in a descriptive manner.
Endpoints	Population size	The model results in population size (and distribution over the classes) over time.
Space	No spatial context	
Time	Days Years	Populations are modelled over several years.
Exposure / effects	Chronic vs. pulse	The simulations consider constant chemical exposure.
Abiotic environment	Photoperiod Different food sources	Photoperiod follows a yearly cycle.

Criteria	Categories	Comments
Biotic environment	Intraspecific competition	Competition through food.
Individuals	Homogeneous	The population is divided into three classes (immatures, females, males), and all individual are identical within a class.
Populations	Others	Population is structured in three classes; each class is modelled with ODEs for mortality, growth (and fecundity).
Calibration	Field data	Basic life history of the mosquitofish is based on field data.
Programming language	STELLA 5.0	Program was implemented into Stella, which is mainly for educational purposes.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	The model did yield population dynamics which were largely consistent with what has been observed in the field.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	The model is presented as Stella diagram, and as the Stella model 'equations'. This makes is very hard to evaluate the model. Most assumptions are not explained.

Assessment

Criteria	Description
Strengths	Specifically tailored for the risk assessment case at hand.
Theoretical uncertainties	The model consists of three classes; within each class the individuals are identical. This makes it impossible to implement more realistic representations of individual behaviour. A constant fraction of the immatures matures in every timestep (independent of their age, as individuals have no age in a stage-structured model).
Empirical uncertainties	This is a descriptive model with many factors that will be difficult to capture in such a simple factor.
Parametric uncertainties	This is a very descriptive model with a number of factors that are put in without sufficient explanation. It is questionable to capture processes such as the number of gravid females into a fixed fraction (or zero, depending on photoperiod and presence of males).
Temporal uncertainties	Photoperiod and food availability follow a fixed pattern over the year. Temperature is not considered.
Conclusions	This model is very descriptive, and appears to include a number of ad hoc factors. Model description and model analysis leave a lot to be desired.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Cabral et al. (2001)	Mosquitofish	Non-ionic surfactant	

1.3.2.5 Spatial ODE Models (Byers and Castle 2005)

Byers and Castle (2005) developed an example of an ODE-based landscape population model to test the efficacy of different integrated pest management (IPM) strategies. Space was modelled as quadratic grid cells representing different agricultural fields. Two scenarios were compared: The first scenario represents local pesticide application on single fields whenever the local pest population size exceeds a certain threshold. The second scenario represents area-wide pesticide application whenever the metapopulation of the area exceeds a certain threshold (coordinated action of farmers). The frequency of applications and the average daily population sizes were recorded from both scenarios and compared. Coordinated action resulted in overall fewer pesticide application because population refugia were precluded from which dispersal could reintroduce insects.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	
Questions / processes	Population recovery Others	Finding the most efficient pesticide application pattern (localized vs. area-wide).
Environmental domain	Terrestrial	Agricultural fields
Taxon specificity	Generic	Generic polyphagous pest species.
Toxicant specificity	Generic	Pesticide
Application	Little-known	
Public availability	Software extension	The model was available as applet on a website which is not accessible anymore.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population size	
Endpoints	Population size Recovery time Other	Average daily population size per field; average daily meta-population size of the area; number of applications to a field.
Space	Grid cell	Landscape with 400 agricultural fields and fallows (grid cells). Each time step, a user-defined fraction of a local population is evenly distributed to the surrounding eight fields (unless some are fallow).
Time	Days	
Exposure / effects	Chronic vs. pulse	Local pulse exposure in a field each time the local population size exceeds a threshold. Or area-wide pulse-exposure in all fields each time the meta-population size exceeds a threshold. Pulse exposure reduces population size to an average fraction defined by the user (0.01 in the demonstration scenario) ± SD.
Abiotic environment	None	No abiotic factors, except for general unsuitability of fallow fields.
Biotic environment	None	No biotic interactions considered.

Criteria	Categories	Comments
Individuals	None	
Populations	Exponential growth	For each field and day, a random growth rate is drawn from a user-defined distribution. No logistic growth used, because application thresholds are set to the exponential growth phase such that populations can hardly approach carrying capacity.
Calibration	-	Calibration of dispersal rate through comparison of results with a previous random-walk individual based model.
Programming language	QuickBASIC 4.5, Java 2.1	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	None	
Sensitivity analysis	Yes	
Uncertainty analysis	No	
Documentation	Scientific publication	Short but concise description of the model.

Assessment

Criteria	Description
Strengths	Simple model with few parameters that may be easily calibrated to different species and toxicant effect levels. The generic nature of the model meets the question it adresses.
Theoretical uncertainties	No toxicity module that relates doses to effects. No conditions considered that may affect the effect size of a pesticide in a population, such as changes in population structure
Empirical uncertainties	No extensive calibration to real world scenarios.
Parametric uncertainties	No effects of biotic or abiotic conditions. No discrimination between toxicant effects on different life stages or sexes and their effects on population growth.
Temporal uncertainties	No seasonal changes in growth rates, no migration of individuals at the edges of the grid ("edge effects").
Conclusions	The model appears useful for scientific research to compare different application scenarios in search of best control strategies. After extensive calibration, the model may be used to compare effects of different exposure patterns on the recovery potential also of non-target species. However, the absolute predictions of the model are very uncertain because numerous potentially relevant aspects such as species interactions and temperature have not been considered.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Byers and Castle (2005)	Generic	Generic	Original publication.
http://www.wcrl.ars.usda.gov/cec/java2/spray.htm			Website

1.3.2.6 Growth Model for Aquatic Plants with 1-Compartment TK (Schmitt et al. 2013)

The model of Schmitt et al. (2013) combines a one-compartment first-order TK module with an ODE-based growth model for *Lemna*. The growth model consists of a simple energy balance: growth is the net result of biomass synthesis and respiration losses. The influence of environmental factors (incl. toxicant stress) are included as factors modifying these two processes. For toxicants, a modified log-logistic equation is used (with a maximum effect level).

General Properties

Criteria	Categories	Comments
Biological level	Population Individual	The model is for *Lemna sp.*, where the distinction between individual and population becomes fuzzy. The model is simple enough to be used in Tier 1.
Model purpose	Regulatory	The model is specifically intended to extrapolate from lab conditions to field conditions.
Questions / processes	Body burden Individual effects Effect propagation Population recovery	The model includes a TK module, considers effects on the 'individual', and propagation to a population of duckweed. Time-varying conditions are allowed, so recovery can easily be followed.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	The model is specific for *Lemna sp.*, but could probably be easily extended to other aquatic plants, as long as they can be modelled as a single homogeneous compartment (e.g., other duckweeds and algae).
Toxicant specificity	Generic	Intended for organic compounds, and compounds that affect the synthesis of biomass (other mechanisms might be included).
Application	Little-known	The model is relatively recent, but built from some well-established principles.
Public availability	Software extension	The model is programmed in R, and the script is included in the publication's supplementary information.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Growth Population density	The model includes a TK module (for body residues) and a growth model (for *Lemna sp.*, the distinction between individual and population growth is largely irrelevant).
Endpoints	Population size Recovery time	The model calculations population size (i.e., total biomass) over a given scenario of (time-varying) environmental conditions.
Space	No spatial context	
Time	Days	Time is considered on a timescale of days to a year.
Exposure / effects	Chronic vs. pulse	Any exposure pattern can be fed into the model.

Criteria	Categories	Comments
	Varying concentrations Repeated exposure	
Abiotic environment	Food limitation Temperature Light	The model includes effects of temperature, light, and nutrients (N and P). All these factors are assumed to act independently by affecting either the synthesis of biomass or the respiration rate.
Biotic environment	Intraspec. competition	A simple form of density dependence was included.
Individuals	Energy budget	The TD model embodies a very simple energy budget: growth is the net result of biomass formation and losses due to respiration. Biomass formation is treated simply as a rate, which is influenced by various environmental factors. The organisms do not have a reserve, which implies that population growth responds immediately to changes in the environment.
Populations	Exponential growth Logistic growth	The distinction between individual and population growth is largely trivial for *Lemna sp*. Under constant conditions, the population will initially grow exponentially, but density dependence will limit the population size.
Calibration	Laboratory data	Parameters for the growth model were established from various literature sources. Toxicant-dependent parameters were either estimated using a QSAR (the BCF) or fitted to the results from toxicity studies.
Programming language	R	R-script available as supplementary information in the paper.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Field data	The growth model was compared to data from field ditches in the Netherlands. The complete model was validated using data on metsulfuron-methyl; it was calibrated on data for constant exposure and subsequently used to predict growth under time-varying exposure. In general, the correspondence with the observed values was very good.
Sensitivity analysis	Yes	Monte Carlo simulation were done for uncertainty/sensitivity analysis, using probability distributions for the most important input parameters. The sensitivity of the parameters depended on the endpoint and the timing of the toxic stress (in a slow- or fast-growth period).
Uncertainty analysis	Yes	
Documentation	Scientific publication	The model is explained quite well in the paper, and the supplementary information provides detailed information on the calibration.

Assessment

Criteria	Description
Strengths	A simple model, based on well-established principles. The model seems to perform very well for its main task: extrapolation from constant conditions in laboratory tests to a time-varying environment.
Theoretical uncertainties	The growth model is a very simple energy budget, with an input (synthesis) and an output (respiration). Environmental factors are assumed to affect these rates, and do so independently (which is questionable). Growth dilution is not included into the TK model, which seems to be an omission. Temperature dependence of the TK needs to be considered. The authors only consider toxicant effects on synthesis, whereas effects might also occur on respiration.
Empirical uncertainties	Standard toxicity tests with *Lemna sp.* contain very little information on parameter values. Therefore, it might be difficult to obtain useful parameter estimates.
Parametric uncertainties	Many different processes are lumped into the growth model into two rate constants; one for synthesis and one for respiration. For example, a nitrogen limitation is modelled as a decrease in the growth rate.
Temporal uncertainties	-
Conclusions	This model seems to be a very useful tool for risk assessment purposes. It is simple enough to interpret the output, and to assess its limitations, yet powerful enough to extrapolate from laboratory conditions to realistic field conditions.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Schmitt et al. (2013)	*Lemna spp.*	Sulfonyl urea herbicide	
Driever et al. (2005)	*Lemna minor*		The basic growth model of Schmitt et al. is based on this paper.
Model applications			
Hommen et al. (2016)	*Lemna spp.*	Sulfonyl urea herbicide	Worked out example of model application in an ERA context.

1.3.2.7 Growth Model for Aquatic Plants with 3-Compartment TK (Heine et al. 2014)

The growth model of Heine et al. (2014) is similar to the one used by Schmitt et al. (2013, see above); a simple energy balance, where growth is the net result of biomass synthesis and respiration losses. The influence of environmental factors (incl. toxicant stress) are included as factors modifying these processes. For toxicants, a log-logistic equation is used to relate internal concentrations to a change in a physiological process. The TK module consists of three compartments (roots, stems, leaves) and is similar to the PBTK models developed for terrestrial plants (see section 1.2.1.4).

General Properties

Criteria	Categories	Comments
Biological level	Individual	The model is simple enough to be used at Tier 1.
Model purpose	Regulatory	Model was developed from a risk-assessment perspective.
Questions / processes	Body burden Individual effects	The model aims to predict growth of *Myriophyllum* as function of environmental conditions. To deal with toxicants, a TK module is added (similar to the Trapp et al models), and a simple TD module (similar to the Schmitt et al model).
Environmental domain	Freshwater	The model is specifically intended for *Myriophyllum*, although it could probably be parameterised for many other rooted aquatic macrophytes.
Taxon specificity	*Myriophyllum sp.*	
Toxicant specificity	Generic	The model is not chemical specific.
Application	Little-known	The model has been developed very recently, but is based on established principles.
Public availability	-	There does not seem to be a publicly available version.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Growth	The model consists of a plant growth model, coupled with a TK module. Internal concentrations affect a parameter in the growth model.
Endpoints	Biomass Recovery time	Model outputs are internal concentrations and plant biomass over time. The authors demonstrate how the model can be used to study recovery of biomass when exposure ceases.
Space	No spatial context	
Time	Dates	Days – one year.
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure	Any exposure pattern can be fed into the model.

Criteria	Categories	Comments
Abiotic environment	Food limitation Temperature Light	The model includes effects of temperature, light, and nutrients (dissolved carbon). All these factors are assumed to act independently by affecting either the synthesis of biomass and/or the respiration rate. Other nutrients (N and P) are assumed to be available ad libitum.
Biotic environment	None	No biotic interactions whatsoever.
Individuals	Energy budget	The TD model embodies a very simple energy budget: growth is the net result of biomass formation and losses due to respiration. Biomass formation is treated simply as a rate, which is influenced by various environmental factors. The organisms do not have a reserve, which implies that population growth responds immediately to changes in the environment. In the fourth thesis chapter (and in Hommen et al), the growth model is extended to include the rhizome, which helps to describe growth in field situations (overwintering, and initial growth in spring).
Populations	-	The model basically described an individual plant, but for these organisms, the distinction between individual and population is rather fuzzy. In the fourth chapter and in Hommen et al, the model is applied as a population model.
Calibration	Laboratory data	Parameter values on *Myriophyllum* were derived from several literature sources. TK parameters are derived using QSARs.
Programming language	Matlab	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data Field data	Some comparisons to laboratory and field biomass data on *Myriophyllum*, and laboratory TK data. In general, the model provided a good description of the observations, although testing was not very elaborate.
Sensitivity analysis	Yes	Effect of temperature and light on the growth rate of the plants. Also potential effect of chemicals on photosynthesis and respiration rates were simulated. Furthermore, effects of all input parameters on predicted effects (fourth chapter).
Uncertainty analysis	No	
Documentation	Scientific publication Website	The model is published in full form in a PhD thesis, from which several publications have been extracted. The model is sufficiently described.

Assessment

Criteria	Description
Strengths	A simple model, based on well-established principles. The model seems to perform very well, although testing has been somewhat limited.

Criteria	Description
Theoretical uncertainties	The growth model is a very simple energy budget, with an input (synthesis) and an output (respiration). Environmental factors are assumed to affect these rates, and do so independently (which is questionable). Growth dilution is not included into the TK model, although it can easily be added. Temperature dependence of the TK needs to be considered.
Empirical uncertainties	Standard toxicity tests with *Myriophyllum* contain very little information on parameter values. Therefore, it might be difficult to obtain useful parameter estimates.
Parametric uncertainties	Many different processes are lumped into the growth model into two rate constants; one for synthesis and one for respiration. For example, the effects of changes in pH and organic carbon are modelled as (independent) decrease in the growth rate in a log-logistic function.
Temporal uncertainties	
Conclusions	This model seems to be a useful tool for risk assessment purposes. It is simple enough to interpret the output, and to assess its limitations, yet powerful enough to extrapolate from laboratory conditions to realistic field conditions.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Heine (2014)	*Myriophyllum spicatum*	Various herbicides	PhD thesis describing the complete model.
Heine et al. (2014)	*Myriophyllum spicatum*	None	Published version of the growth model.
Heine et al. (2015)	*Myriophyllum spicatum*	Various herbicides	Published version of the TK module.
Model applications			
Hommen et al. (2016)	*Myriophyllum spicatum*	Sulfonyl urea herbicide	Worked out example of model application in an ERA context.

1.3.2.8 DEB for Unicellulars (Hanegraaf and Muller 2001)

DEB theory is a theory for all of life, including unicellular organisms (see section 1.2.2.2). A population of these organisms is often treated in DEB applications as a superindividual with specific shape coefficient (V1-morph). Here an example of Hanegraaf and Muller (2001) has been reviewed.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	
Questions / processes	Effect propagation Others	Some applications are from a (microbial) ecology perspective, but also some from an ecotoxicological perspective.
Environmental domain	Generic	Applied to bacteria and algae (and corals, in modified form).
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	A number of publications use variations of this model, although it will be little known in ecotoxicology and regulatory settings.
Public availability	Software extension	There is probably a DEBtool package (under Matlab) that can do some of these calculations, but user-friendliness is limited.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality Growth	
Endpoints	Population size	Endpoint is the number or biomass of cells in the population.
Space	No spatial context	
Time	Dates	
Exposure / effects	Chronic vs. pulse exposure	The cases studied work with constant exposure, but time-variable stress could be easily accommodated.
Abiotic environment	Food limitation Temperature	
Biotic environment	-	
Individuals	Energy budget Stochastic	Stochastic for cell death.
Populations	Exponential growth	Population is described as a super-individual. Under constant conditions, this yields exponential growth, and under limiting conditions, e.g., logistic growth may result.
Calibration	Laboratory data	

Criteria	Categories	Comments
Programming language	Various (probably most in Matlab)	No user-friendly software available.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	Multiple calibrations on data sets.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Detailed representation of energetics in unicellulars.
Theoretical uncertainties	
Empirical uncertainties	
Parametric uncertainties	
Temporal uncertainties	
Conclusions	Standard data are generally insufficient to calibrate these models, and detailed effects modelling of unicellulars may be less relevant for risk assessment.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman (2000)	Generic	None	General book explaining DEB theory in all its aspects.
Hanegraaf and Muller (2001)	Bacteria	None	Basic model for unicellulars.
Model applications			
Muller et al. (2009)	Corals	None	Including the symbiontic interaction between coral and algal symbiont.
Klanjscek et al. (2012)	Bacteria (*Pseudomonas*)	Cadmium	Extension of the bacterial model with toxic effects.
Klanjscek et al. (2013)	Bacteria (*Pseudomonas*)	Cd and CdSe nanoparticles	
Eynaud et al. (2011)	Corals	Radiation	
Muller and Nisbet (2014)	Algae (*Coccolithophores*)	Ocean acidification	

1.3.2.9 DEBtox for Unicellulars (Kooijman et al. 1996)

The DEBtox algae model of Kooijman et al. (1996) was part of the DEBtox software for the analysis of standard toxicity data. It is a simple extension of the exponential growth model to account for three potential mechanisms of toxicity: increasing costs for growth, hazard rate (increasing probability for a cell to die), and adaptation (as previous one, but for a limited time period). The model ignores toxico-kinetics (instantaneous steady state).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Regulatory	
Questions / processes	Effect propagation	
Environmental domain	Generic	Applied to algae, but in principle, this model approach should work well with many small organisms (e.g., unicellulars).
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Little-known	Exponential growth of algae is well established, but the various options to include toxicant effects in the equation are not.
Public availability	-	Used to be part of the DEBtox software that is not updated anymore. Easy to program in many software environments.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Mortality Growth	
Endpoints	Population size	Endpoint is the number of algal cells in the population.
Space	No spatial context	
Time	Dates	
Exposure / effects	Chronic vs. pulse exposure	The model is set up to analyse data from standard tests, and thus constant exposure. Time-varying exposure should be possible when effects are fully reversible.
Abiotic environment	-	Effects of temperature can easily be added.
Biotic environment	-	
Individuals	Stochastic	Stochastic for cell death.
Populations	Exponential growth	The model is based on exponential growth, although some mechanisms of action will yield a deviation from exponential.

Criteria	Categories	Comments
Calibration	Laboratory data	Model intended to be calibrated from standard algae tests.
Programming language	-	No implementation seems to be available anymore. However, the model is so simple, it can easily be programmed in many software environments (incl. Excel).

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	Multiple calibrations on data sets.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Very simple model that is able to work with data from standard algal toxicity test. The model uses all test data to derive the model parameters (such as a no-effect concentration).
Theoretical uncertainties	All physiology of algae is reduced into an exponential growth rate, and toxicants affect either this growth rate or the probability to die.
Empirical uncertainties	
Parametric uncertainties	
Temporal uncertainties	
Conclusions	This very simple model can be used a tool to analyse toxicity data, using all data from the test in a single analysis. Its simplicity may prevent useful extrapolations beyond the conditions of the test.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman et al. (1996)	Algae	Range of compounds and two mixtures.	Presentation of the basic model and demonstration on several data sets.
Model applications			
Pablos et al. (1998)	Trout red blood cells	Chlorophenol	
Urrestarazu Ramos et al. (1999)	Algae	11 polar narcotics	
Arzul et al. (2006)	4 algae (freshwater and marine)	Carbofuran and isoproturon	
Miller et al. (2010)	4 species of algae	Nanoparticles	Extension of the model to include a more detailed mechanism of action.

1.3.2.10 **Euler-Lotka Equation with DEBtox (Jager et al. 2004)**

DEBtox has been described in section 1.2.2.3. As DEBtox describes survival and reproduction over the life cycle (and the effects of stressors on these traits), the output can be easily linked to the Euler-Lotka equation (here in its continuous form). This equation integrates survival and reproduction over the entire life cycle to generate the intrinsic rate of population increase. An example of Jager et al. (2004) has been reviewed here.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Individual effects Effect propagation	
Environmental domain	Generic	
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Both DEBtox and the Euler-Lotka equation are well-established in Science. Their combination is not too common.
Public availability	Software extension	Several Matlab implementations.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Effects on individuals are treated in the DEBtox module, and they are propagated to a population growth rate using the Euler-Lotka equation.
Endpoints	Survival Reproduction Population size	Endpoint is the exponential growth rate of the population under constant conditions.
Space	No spatial context	
Time	Dates	The EL calculation requires survival and reproduction over one life cycle, and translates this into a population growth rate.
Exposure / effects	Chronic vs. pulse exposure	Only chronic, constant, exposure. This limitation is caused by the EL calculation, and not by the DEBtox module.
Abiotic environment	Food limitation Temperature	Different food levels and temperatures can be used in the calculation, as long as they remain constant over time.
Biotic environment	-	No biotic factors, apart from the individual's energy budget.
Individuals	Stochastic Energy-budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	Exponential growth	Exponential growth rate is calculated.

Criteria	Categories	Comments
Calibration	Laboratory data	DEBtox models are calibrated for each case; the EL calculation requires no additional parameters.
Programming language	Matlab	Several Matlab implementations are available. For example, within the BYOM framework: http://www.debtox.info/byom.html

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data No independent data	The DEBtox module has been 'validated', but the EL calculation has not (it is merely book-keeping).
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication Website	Extensive documentation is available for the DEBtox module (see that sheet); the EL has a useful Wikipedia entry.

Assessment

Criteria	Description
Strengths	The DEBtox model is a well-established TKTD model for individuals. The EL calculation is extremely simple and easy to interpret due to the total lack of ecological complexity. The intrinsic rate of increase is useful as a fitness measure for the population under stress.
Theoretical uncertainties	The DEBtox module is based on a rigorous simplification of animal energetics over the entire life cycle. Inevitably, details on life history will be lost. The EL calculation assumes a constant environment over many generations, which is unrealistic.
Empirical uncertainties	Many species, or data sets, require adaptations to the DEBtox model; often difficult to find a unique mechanism of action of a chemical; no applications to birds and mammals.
Parametric uncertainties	Often difficult to identify all DEBtox parameters from standard data sets. The EL calculation does not require additional parameters.
Temporal uncertainties	The EL calculation is based on a constant environment over many generations. This is unrealistic.
Conclusions	Euler-Lotka is a simple way to extend the DEBtox TKTD model to the population level. The resulting growth rate is easy to interpret, and should be seen as a measure of the fitness of the population rather than a realistic representation of population dynamics.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Jager et al. (2004)	Springtails	Cadmium, tributyltin	First complete example of DEBtox linked to the Euler-Lotka equation. Extrapolation to different food levels.
		Model applications	
Alda Álvarez et al. (2005)	Nematodes (*C. elegans*)	Cadmium	Extrapolation to different food levels.
Alda Álvarez et al. (2006a)	Nematodes (*A. nanus*)	Cadmium, pentachlorobenzene, carbendazim	Extrapolation to different temperatures.
Jager et al. (2007)	Springtails	Chlorpyrifos	Identification of two modes of action.
Muller et al. (2010)	*Daphnia*	Tetradifon, pyridine	Slightly different DEB-based model, extrapolation to different food levels.
Jager and Klok (2010)	Earthworms	Copper	Comparing different DEB models and population approaches.
Jager and Zimmer (2012)	*Daphnia*	Fluoranthene	Population growth with 95% credible intervals.

1.3.2.11 **Euler-Lotka Equation with kmDEB (Kooijman and Metz 1984)**

Kooijman-Metz DEB is a predecessor of DEB theory, and bares a close resemblance to DEBtox. It is based on a simplified energy budget. In the example of Kooijman and Metz (1984b), the output in terms of survival and reproduction is linked to the continuous form of the Euler-Lotka equation. In comparison to DEBtox, kmDEB has no reserve, no maturity maintenance, and no TK considerations (although survival effects are time dependent).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Individual effects Effect propagation	
Environmental domain	Generic	
Taxon specificity	Generic	The model is generic although only application to *Daphnia* is presented in the original paper.
Toxicant specificity	Generic	
Application	Little-known	kmDEB is a predecessor of DEB theory, and closely resembles DEBtox. Here, it is linked to the Euler-Lotka equation to calculate population growth rate.
Public availability	-	No implementations available (original was likely programmed in APL).

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Effects on individuals are treated in the kmDEB module, and they are propagated to a population growth rate using the Euler-Lotka equation.
Endpoints	Survival Reproduction Population size	Endpoint is the exponential growth rate of the population under constant conditions.
Space	No spatial context	
Time	Dates	The EL calculation requires survival and reproduction over one life cycle, and translates this into a population growth rate.
Exposure / effects	Constant exposure	Only chronic, constant, exposure. This limitation is caused by the EL calculation, and partly by the kmDEB module which lacks a TK module.
Abiotic environment	Food limitation	Different food levels can be used in the calculation, as long as they remain constant over time.
Biotic environment	-	No biotic factors, apart from the individual's energy budget.

Criteria	Categories	Comments
Individuals	Energy budget Stochastic	Stochastic for survival (individual tolerance model), energy budget for growth and reproduction.
Populations	Exponential growth	Exponential growth rate is calculated.
Calibration	Laboratory data	Model is fitted to data; the population predictions do not require additional parameters.
Programming language	APL?	There is no up-to-date implementation of this model, though it would be easy to programme in many software applications.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation attempts known. Model is based on first principles and describes life history patterns quite well.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	EL has a useful Wikipedia entry.

Assessment

Criteria	Description
Strengths	The EL calculation is extremely simple and easy to interpret due to the total lack of ecological complexity. The intrinsic rate of increase is useful as a fitness measure for the population under stress.
Theoretical uncertainties	The kmDEB model is not used anymore, as it has been replaced by DEBtox. The EL calculation assumes a constant environment over many generations, which is unrealistic.
Empirical uncertainties	
Parametric uncertainties	
Temporal uncertainties	The EL calculation is based on a constant environment over many generations. This is unrealistic.
Conclusions	Euler-Lotka is a simple way to extend the DEB-based TKTD model to the population level. The kmDEB model is outdated and has been replaced by DEBtox.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kooijman and Metz (1984b)	Generic, *Daphnia*		
Model applications			
Jager and Klok (2010)	Earthworms	Copper	Comparison of different DEB models and simple population approaches.

1.3.2.12 Partial Differential Equation (PDE) Model with Energy Budget (Hallam et al. 1990)

The model of Hallam et al. (1990a) applies an energy budget for the individuals, which has some similarities to DEB but is less well established. The population is modelled as a partial differential equation (PDE; here the McKendrick-von Foerster equation) leading to an age-structured population model (in essence: cohorts of individuals are followed over time, and all individuals with the same age have the same properties unless they belong to different ecotypes). The model has (as far as we could see) not been compared to measured data at the individual level nor at the population level. Differences within the species are included as different 'ecotypes'; each ecotype has its own parameters and requires a separate PDE.

General Properties

Criteria	Categories	Comments
Biological level	Individual Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Individual level effects Effect propagation	
Environmental domain	Freshwater	The model was developed for *Daphnia magna*, but may be adapted to other species.
Taxon specificity	Taxon-specific	
Toxicant specificity	Generic	Used to simulate effects on narcotic chemicals.
Application	Little-known	
Public availability	Software extension	

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Mortality Growth Population density	
Endpoints	Population size	
Space	No spatial context	
Time	Dates	Time is included continuous.
Exposure / effects	Chronic vs. pulse	Some versions of the model include a one-compartment TK module.
Abiotic environment	Food limitation Temperature	One version includes dissolved oxygen as an additional environmental factor.
Biotic environment	Intraspec. competition	Competition between 27 different 'ecotypes' of *Daphnia*.

Criteria	Categories	Comments
Individuals	Energy budget	Individuals are described by an energy budget (different from DEB theory). In the population model, all individual with the same age have the same properties, although several ecotypes are modelled as separate species.
Populations	Other	Age-structured, continuous time, in a PDE.
Calibration	Laboratory data	Literature data are used to derive a set of parameters for the modelled species.
Programming language		Not reported, no version seems to be available.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No comparisons to measured data seem to be made.
Sensitivity analysis	Yes	The influence of individual-level parameters on several individual-level endpoints.
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Detailed treatment of lipid dynamics in Daphnia, including moulting behaviour and the discrete deposition of eggs in the brood pouch. The population approach allows a useful representation of individual traits over the life cycle for each individual.
Theoretical uncertainties	It is not clear to what extent the energy-budget model is able to describe the life-history traits of individuals (under time-varying environmental and exposure conditions). The population approach is limited in that all individuals of the same age (and ecotype) have the same properties.
Empirical uncertainties	
Parametric uncertainties	The model requires a rather large number of parameters, which have been set in various ways.
Temporal uncertainties	Environment is taken as constant.
Conclusions	Model cannot be used for risk assessment, at least until it has been established that the individual model provides a reasonable representation of individual life-history traits (also under toxicant exposure).

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Hallam et al. (1990b)	*Daphnia magna*	None	Description of the energy-budget model for the individuals.
Hallam et al. (1990a)	*Daphnia magna*	Narcotic chemicals	Population approach added, and toxicant effects on mortality (general chemical stress).
Model applications			
Hallam et al. (1993)	*Daphnia magna*	Narcotic chemicals	Extension to effects on growth (general chemical stress).
Hallam et al. (2000)	Fish	None	Simulation of fish population dynamics.
Koh et al. (1997)	*Daphnia magna*	Narcotic chemicals	General chemical stress, combined with effects of temperature and dissolved oxygen.

1.3.2.13 **Delay-Differential Equation (DDE) Models (Brown et al. 2003)**

Delay-differential equations (DDEs) can be used to model a population structured in age classes. Within a class, all individuals are identical, and each class is followed over continuous time (in practice: with 1 -day time step). In the reviewed example of Brown et al. (2003), the chemical affects life-history traits directly (no TKTD model for the individuals).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Regulatory	Authors intended this paper as a contribution to the risk assessment for endocrine disruptors in fish.
Questions / processes	Effect propagation Population recovery	Effects on individual life-history traits are propagated to the population level, and the model is also used to simulate recovery after exposure.
Environmental domain	Freshwater	
Taxon specificity	Generic	Freshwater fish, but could be modified to capture the life cycle of other species (with appropriate parameterisation).
Toxicant specificity	Generic	The model is applied to endocrine disruptors, but may be applied more generally. The model does not include toxicokinetic aspects, but assumes a direct and constant link between the external concentration and the life-history traits.
Application	Little-known	This specific model does not seem to have been applied a lot, but similar DDE models have been regularly applied in population ecology.
Public availability	-	There does not seem to be a publicly available version.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Age class	Each class is modelled by considering recruitment into the class (from spawning in the first class), maturation to the next class, and mortality.
Endpoints	Population size Recovery time	The model follows population size over time, and can predict the probability of extinction and the time needed to recover after a disturbance.
Space	No spatial context	
Time	Days	Each class is modelled with a time step of one day, and the population is modelled over 25 years.
Exposure / effects	Chronic vs. pulse	The simulations consider constant chemical exposure for several years, followed by a recovery in a clean situation.

Criteria	Categories	Comments
Abiotic environment	None	No description of the environment. Vital rates are kept constant, and are only influenced by competition and chemical exposure.
Biotic environment	Intraspecific competition	Density dependence through competition with conspecifics. Different form of density dependence is used for different fish species.
Individuals	Homogeneous	The population is divided into age classes, and all individual are identical within a class. Stochasticity is included on the fecundity rate only.
Populations	Age classes	Population is structured in age classes.
Calibration	Laboratory data Field data	Calibration with various freshwater fish. Basic life history is mostly based on field data. Effect of chemicals on vital rates is based on laboratory data (in some cases for other species of fish).
Programming language		Unknown

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation.
Sensitivity analysis	Yes	Different model runs with different vital rates were performed to provide a qualitative impression of the sensitivity.
Uncertainty analysis	Yes	Extinction risk was calculated from Monte Carlo simulations, using distributions for the vital rates.
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Relatively simple way to model dynamics of fish populations, including recovery and extinction risk. However, it is unclear what this specific model implementation (as DDEs) adds to simpler Leslie matrix models.
Theoretical uncertainties	The model has no representation of individuals, and hence no toxicokinetics or toxicodynamics. Chemical stress has an immediate and constant effect on the vital rates. Structuring the population in age classes makes it difficult to include more realistic representation of the individuals (just as with matrix models).
Empirical uncertainties	
Parametric uncertainties	All toxicokinetics and -dynamics reduced to an immediate and constant on the vital rates. Predation is included as a mortality rate, no modelling of the food or competition with other species.
Temporal uncertainties	The environmental factors (e.g., food and temperature) are not explicitly included in the model and hence are taken as constant.

Criteria	Description
Conclusions	The model is relatively simple to use, but seems to offer few benefits over even simpler approaches (e.g., Leslie matrix). For a more realistic inclusion of individual-level behaviour, an IBM approach would be far superior. This particular model does not seem to be applied much, although the concept of DDEs has been more widely applied in population ecology.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Brown et al. (2003)	Brook trout and fathead minnow	Nonylphenol and methoxychlor	First model description.
Model applications			
Brown et al. (2005)	Perch	Nonylphenol and ethinylestradiol	Studies the impacts of the chemicals in the population before and after a disease, which affected the vital rates.

1.3.3 Individual-Based Models (IBMs)

1.3.3.1 Connected Individual and Population Models for Seals (Hickie et al. 2005)

The model of Hickie et al. (2005) for the bioaccumulation of persistent organic pollutants (POPs) in seals provides an example of how sub-models for various life stages can be connected to a life table population model. Each life stage is thus modelled as a single "super-individual", leading over to the true individual-based modelling approaches where every individual can have different properties and state variables. The model predicts POP burden and thus is actually only an exposure but not an effect model.

The individual-based model (IB) of Hickie et al. (2005) calculates the accumulation of POPs over the entire life of an individual, taking into account the animal's complete life history. The IB model tracks the growth and energetics of an individual, as well as contaminant accumulation, disposition between two compartments (blubber and "core") and elimination for an individual seal (male or female) on a daily basis from weaning (~40 days of age) until death (~30 years of age). It includes an additional subroutine for gestation, birth and nursing which can be invoked for any year after a female reaches maturity. Contaminant absorption from the diet is the only uptake pathway considered and is a function of the concentration in prey, daily ration and the absorption efficiency. Chemical-specific elimination via faeces, biotransformation and for some chemicals via respiration is described by a single first-order elimination rate constant (k_e).

The population-based (PB) model uses energy and contaminant flux budgets summarized from the IB model and quantifies the changes in contaminant levels throughout the population over several generations. Specifically, the PB model combines a population life table consisting of 30 age-classes (years) for each sex, with sexually mature females being further subdivided into pregnant/nursing and resting categories. The PB model proceeds in yearly time steps and runs the IB sub-model for each subpopulation at each time step. Only the contaminant concentration in food changes across years in the PB model, so that cohorts can differ in their body burden due to the history of dietary exposure.

General Properties

Criteria	Categories	Comments
Biological level	Population	Aim was to hindcast temporal trends and to predict potential future trends in contaminant loading of juvenile seals.
Model purpose	Scientific	
Questions / processes	Effect propagation	Aim was to simulate the bioaccumulation of persistent organic pollutants (POPs) over the lifetime of ringed seals (*Phoca hispida)* including maternal transfers to progeny and to account for the effects of age, growth, body condition and sex.
Environmental domain	Marine	
Taxon specificity	Taxon-specific	
Toxicant specificity	Toxicant specific	Five persistent organic pollutants (POPs) such as DDT and PCBs were used.
Application	Established in science	30 / 19 citations google scholar/ Web of Science.
Public availability	No	

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Mortality Growth Reproduction	Life history parameters modelled with a range of functions: Gompertz growth curve, allometric length-mass regression, energy intake, metabolics and contaminant kinetics.
Endpoints	Body burden	Contaminant trends.
Space	No spatial context	
Time	Dates	IB model on daily basis; PB model in annual transitions.
Exposure / effects	Varying concentrations	Contaminant absorption from the diet is the only uptake pathway considered and is a function of the concentration in prey, daily ration and the absorption efficiency. Chemical-specific elimination via faeces, biotransformation and for some chemicals via respiration is described by a single first-order elimination rate constant (k_e).
Abiotic environment		Not considered; important is the contaminant load in prey influencing the intake rate.
Biotic environment		Not considered.
Individuals	Energy budget	No stochasticity; the IBM structure is used to program more complex subroutines that cannot be implemented in the age- and stage-based matrix model structure.
Populations	Other	Not considered.
Calibration	Field data	A dataset with 50 seals (29 males, 21 females) collected from Arviat, Nunavut in 1993, was used to assess the first-order elimination rate constant (k_e).
Programming language	-	Unknown.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Data for male and female seals from Holman Island from 1972 to 2001 (Addison and Smith 1974; Hoekstra, unpublished) and for females from Arctic Bay (Ikpiarjuk) Nunavut from 1975 to 2000 (Muir et al. 2001).
Sensitivity analysis	-	
Uncertainty analysis	Yes	Simple uncertainty analyses for PCB 180 in males with 10% variation from baseline values for combinations of asymptotic length (A), blubber volume fraction, metabolic rate activity factor (AE) and k_e gave results that showed consistency with findings from studies that concentrations of POPs in blubber should be negatively correlated with body condition.
Documentation	Scientific publication	No specific documentation; methods section in publication.

Assessment

Criteria	Description
Strengths	Model based on first principles (energetics) and calibrated and validated with field data.
Theoretical uncertainties	
Empirical uncertainties	No focus on population effects and varying environments.
Parametric uncertainties	Very specific parameters for seals; may not be easy to obtain for other species.
Temporal uncertainties	PB model integrated over yearly time steps which might lead to uncertainties in outcome.
Conclusions	Well-designed IBM for focal species, difficult to adjust to other taxa.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Hickie et al. (2000)	St. Lawrence beluga population	A variety of POPs (select PCB congeners, DDTs, chlordanes, and HCHs)	This is the original model description for the individual- and population-based models.
Hickie et al. (2005)	Arctic ringed seal (*Phoca hispida*)	Persistent organic pollutants (POP)	The main publication reviewed here.

1.3.3.2 **BEEHAVE (Becher et al. 2014)**

BEEHAVE, first described in Becher et al. (2014), is a model under development for the risk assessment of pesticides on honeybees. The model pools life stages in the hive, and groups of 100 individuals ("superindividuals") outside the hive that are considered to behave similarly. This model concept is comparable to that of Hickie et al. (2005, see above). In the model, a variable number of eggs laid by the queen develop into larvae, pupae, and in-hive worker bees or drones. Worker bees turn into foragers at a specific age dependent on the bee population structure and the food storage in the hive. Workers die after a specific age or flown distance. Successful foragers raise the honey and pollen storage, in-hive bees consume food. Foraging behaviour of workers is guided by energy efficiency and flight duration. Weather, location and food supply of a patch (field) affect the decision. *Varroa destructor* mites randomly invade larval cells, where they infect and get infected from bee pupae with a DWV-virus. Reproduction of the mites depends on the bee density in an invaded cell.

General Properties

Criteria	Categories	Comments
Biological level	Population (extended)	Population model of honey bees with consideration of a pathogen and its vector population.
Model purpose	Scientific / Regulatory	
Questions / processes	Population recovery	Predict population dynamics and failure in the presence of multiple stressors including *Varroa* mites acting as virus vectors, impaired foraging behaviour, changes in landscape structure and dynamics, and pesticides.
Environmental domain	Terrestrial	*Apis melifera* (honey bee)
Taxon specificity	Taxon-specific	
Toxicant specificity	Generic	
Application	Applied in studies	
Public availability	Stand-alone programme	www.beehave-model.net

Variables and Parameters

Criteria	Categories	Comments
Entities	Age cohorts Individual groups Population	In-hive bee population represented by age cohorts and further distinguished by sex and exposure to mites and viruses. Foraging bees modelled as "super-individuals" (group of 100 individuals with identical behaviour). *Varoa* mites represented by the population size of uninfected and infected individuals attached to bees, viruses modelled as transmission rates. The hive is represented by honey and pollen storage, the queen by reproduction rate.
Endpoints	Extinction	Survival or failure of colonies. Colonies were considered to die when the population size fell under a threshold of 4,000 individuals.

Criteria	Categories	Comments
Space	Continuous Landscape (km²)	Implicit description of space: Dynamic flower patches are characterized by probabilities to be found by foragers according to size and distance from the hive. Maps of real landscapes can be imported and transferred to probability values.
Time	Days, hours	Bees, mites, colony and landscape change in daily steps. The foraging module is executed daily and contains a varying number of foraging trips (in minutes) depending on the available food sources, demand and weather.
Exposure / effects	Not specified	No exposure module included. Mortality rates of foraging bees (extension: also of different life stages) can be increased and egg production rate decreased by the user to simulate the effect of pesticides contamination.
Abiotic environment	Food limitation Temperature Light	Flower patches as food sources, weather conditions affect the daily foraging period.
Biotic environment	Parasitism	Pathogen vector (*Varroa* mites) population; viruses modelled as infection rates.
Individuals	Stochastic	Squadrons of 100 foraging bees make behavioural decisions based on energy efficiency. *Varoa* mites randomly move between larval cells.
Populations	Age cohorts	Age-cohorts of in-hive bees, super-individuals of foraging bees.
Calibration	Laboratory data Mesocosm data	*Apis melifera*, *Varroa destructor*
Programming language		NetLogo 4.1.3

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data (*Apis melifera*)	No statistics, only visual comparison of trends, because underlying environmental conditions of field observations were not available. Output for the emerging properties colony dynamics, life span and age of becoming a forager from 10 model runs showed similar trends to field data for 1 year. Predicted forager Behaviour coincided with results from a foraging experiment. Predicted colony failure after 3 -4 years of Varroa infection coincided with field observations.
Sensitivity analysis	Yes	61 parameters tested individually (no interactions), output from 10 replicate simulations over 3 years was studied. Generally low sensitivity due to compensatory feedback mechanisms in the model. Most sensitive parameters related to mortality, energy influx, colony growth.
Uncertainty analysis	Yes	95 % CI from repeated simulations reported.
Documentation	Website Scientific publication	Detailed description in the first publication following the ODD protocol. Comprehensive help functions on the website.

Assessment

Criteria	Description
Strengths	Detailed consideration of energy flux and interactions with parasite vectors in bee colonies, and of foraging behaviour under varying weather conditions and food supply.
Theoretical uncertainties	No interactions of different stressors within an individual. Limited dynamic task allocation between individuals, no temperature regulation within the hive.
Empirical uncertainties	Most parameters calibrated based on estimations rather than exact measurements.
Parametric uncertainties	Mortality of foraging bees based on simple probabilities, not based on simulations of real stressors such as pesticides. Therefore, no sublethal effects of pesticides considered. Only a single virus and vector species.
Temporal uncertainties	
Conclusions	Reasonable compromise of simplicity and realism for the dynamics of bee colonies. More rigorous validation of model predictions necessary. Missing of the explicit modelling of pesticide effects currently limits applicability for risk assessment.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Becher et al. (2014)	*Apis melifera*, *Varroa destructor*	None	Original model description.
Rumkee et al. (2015)	*Apis melifera*, *Varroa destructor*	Generic	Extension to increase mortality of different life stages through pesticide exposure. Applied in a highly simplified scenario, no tests with real data. Increased adult mortality had a higher effect on colony failure than increased larval mortality or reduced reproduction. Calculation of LIS50 (imposed stress to kill a colony in 50 % of simulations).
Model applications			
Horn et al. (2016)	*Apis melifera*	None	Application to study the effects of spatially and timely gaps in food supply in simplified scenarios. Cascading effects of food gaps drive colonies to extinction.
EFSA PPR (2015b)	*Apis melifera*, *Varroa destructor*	Generic	Stepwise evaluation of BEEHAVE for regulatory risk assessment: Good modelling of colony dynamics, but missing pesticide module. Underestimation of the effects of virus infection, missing interactions of multiple stressors.
McMahon et al. (2016)	*Apis melifera*, *Varroa destructor*	None	Application to support experimental and field studies on the effects of a new DWV-virus strain on honey bee colonies.

1.3.3.3 **IDamP (Preuss et al. 2009)**

The individual-based *Daphnia magna* population model (IDamP, Preuss et al. 2009a) is an example of a simple "true" individual-based model (IBM) in which every individual is simulated distinctly and can have unique parameter values and state variables. The model is non-spatial and has been coupled in Preuss et al. (2010a) with a simple dose-response module for the toxic effects of pesticides on individuals. The model differs from other IBMs by the consideration of crowding as a density-dependent reduction in growth and survival (if experienced as subadult) that is independent from food limitation, which generally decreases survival. It was designed to work with data from standard ecotoxicological tests. In the model, the development, survival and reproduction of a daphnid depends on both the food availability and the population density during each time step, calibrated with data from a life table response experiment (LTRE). Food availability is modelled through a separate algae population that is represented by simple differential equations. 12 parameters describe the crowding phenomena, in total the model uses 37 parameters. The toxicant can cause acute mortality and reduced reproduction in the model. The model simulates populations of two interacting species (*Daphnia* and food) and has later been extended to the simulation of a second competing daphnid population and predation at a constant rate, so that it may be considered also as a community model. However, the focus is clearly on the simulation of the *Daphnia* population within its environmental context, therefore IDamP was considered as an extended population model here.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific Regulatory	
Questions / processes	Effect propagation Population recovery	The original model aims at demonstrating the effect of crowding. Extensions combine the model with toxicological data of chemicals to predict the effect propagation from the individual to the population under crowded conditions.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	*Daphnia magna*, *Desmodesmus subspicatus* (alga serving as food)
Toxicant specificity	Generic	No toxicant exposure included in the original model. Later the model was applied e. g. to 3,4-dichloroaniline.
Application	Established in science	
Public availability	Source code	

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Individuals	
Endpoints	Population size Population structure Survival	When a toxicant was included, the concentration-dependent mean population size, the population size structure in equilibrium, and the extinction probability after 150 days of exposure were reported.
Space	No spatial context	Food and daphnids are considered to be homogeneously distributed in the water (no explicit spatial context). Crowding is considered based on the population density (volume of all individuals / volume of vessel) and calibrated to vessels of a few liters.
Time	Das Hours	Daphnia is modelled in time steps of days, the algal population is modelled in hours.
Exposure / effects	Chronic vs. pulse Varying concentrations	Acute mortality and decreased reproduction depend on the exposure concentration, as calibrated with dose-response curves from standard acute and reproduction tests.
Abiotic environment	Food limitation	The original model was calibrated to a fixed water temperature of 20°C which cannot be changed.
Biotic environment	Intraspec. competition Crowding Predation Interspec. competition	Algae serving as food are modelled as a prey population. Rates of algal immigration and emigration can be additionally specified by the user. Interspecific competition with a second *Daphnia* population available in an extension.
Individuals	Stochastic	Each individual is characterized by a specific filtration rate, body length, duration of juvenile development and lifetime, that is chosen from a normal distribution at birth (no heredity of trait characteristics). Acute sensitivity is chosen from a log-normal distribution calibrated with acute dose-response curves.
Populations	Individual-based	Algal population modelled with differential equations.
Calibration	Laboratory data	Original model: Data from life cycle experiments with *D. magna* at flow-through conditions with different levels of algae concentrations. Variability in these data were used for calibration of the stochastic functions that generate variation in the traits between the individuals.
Programming language	Delphi	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	Without toxicant: Predictions of reproduction (individual level) were tested with 21-day reproduction tests at semi-batch conditions with different food conditions. Predictions deviated by < 2 % from the observations. Predictions of population dynamics (population level) were tested with population experiments at different food levels under flow-through and static conditions. Good prediction of the abundance over time (deviations < 10 - 25 %) and moderate prediction of the size structure, except for low food conditions. With exposure to 3,4-dichloroaniline, the equilibrium population density was 39 % higher as compared to test observations, and the growth rate deviated by 26 %.
Sensitivity analysis	No	
Uncertainty analysis	Yes	1000 Monte Carlo runs to produce 95 % CI for the model predictions.
Documentation	Scientific publication Website	The model is explained in a well-structured publication according to the ODD protocol.

Assessment

Criteria	Description
Strengths	The model considers many relevant mechanisms that affect the population dynamics of aquatic macroinvertebrates. It has been tested with a number of independent data and has been shown to predict the effect of different food levels and toxicant concentrations on the individual development, and on the population dynamics of *Daphnia magna* with moderate to good precision under controlled conditions.
Theoretical uncertainties	The model does not consider adaptation in the behavior (filtering, switch from r- to K-strategy in reproduction) to different food conditions and population densities. This may be a reason for the imprecise predictions of the population structure. The mechanisms through which crowding acts are not understood.
Empirical uncertainties	The required LTRE input data are only valid for the controlled test conditions, LTRE data under various field conditions are difficult to obtain.
Parametric uncertainties	Few abiotic stressors and no migration considered. Individuals may respond differently to crowding (r- and K-strategists within a population), therefore various trait characteristics may be correlated and heritable, but are modelled as independent and stochastic. No sublethal effects except for reproduction considered and no prediction of mixture effects.
Temporal uncertainties	The original model was calibrated to fixed laboratory conditions in terms of water quality etc. Food (and in the extended version temperature) are the only environmental conditions that can be varied. No toxicodynamics (only immediate effects assumed). In the extended version, predation is implemented as a constant rate (no population dynamics of a predator).

Criteria	Description
Conclusions	The model is specific for *Daphnia magn*a and simulates the population dynamics and the effects of toxicants well under laboratory test conditions. The model is under continuous enhancement to extend the range of possible scenarios, but is generally limited to the simulation of populations under artificial conditions.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Preuss et al. (2009a)	*Daphnia magna*	None	Original model without toxicity module.
Preuss et al. (2010a)	*Daphnia magna*	Dichloroaniline	Introduction of the toxicity module, tested with constant exposure only.
http://www.bio5.rwth-aachen.de/index.php/forschung/modellierung-und-simulations/27-idamp-model			Website
Model applications			
Gergs et al. (2013b)	*Daphnia magna*, *Notonecta maculata* (predator)	p353-nonylphenol	Extension includes pulsed exposure combined with predation by an insensitive predator or competitor. Kairomone-induced reduction in clutch size considered. Results tested with laboratory microcosm data. Species interactions reduced the population resilience even at low concentrations that had otherwise no detectable effect in the model.
Gabsi et al. (2014a)	*Daphnia magna*	None	Extension that links the offspring size to maternal traits.
Gabsi et al. (2014b)	*Daphnia magna*	Dispersogen A	Compares predictions of toxicant effects on population size and extinction risk based on individual-level test data by IDamP and by traditional growth models. Results highly dependent on assumptions about individual-level effects; transmission of effects on F1 had to be integrated.
Gabsi and Preuss (2014)	*Daphnia magna*, *Chaoborus crystallinus* (predator)	Hypothetical generic toxicant	Extension allows to vary the temperature and includes predation and interspecific competition by an insensitive antagonist. Analysis of the recovery time after pulse exposure to LC30 - LC80 alone and together with either constant predation by *Chaoborus* or a hypothetical competing *Daphnia* species. The model predicted interactive effects of food level, temperature, concentration and species interactions. Results were not tested with real data, but fitted better to some case studies from literature than traditional exponential growth models.

Citations	Taxa	Chemicals	Comments
Gabsi et al. (2014d)	*Daphnia magna*	Hypothetical generic toxicant	Extension includes predation and interspecific competition by an insensitive antagonistic population, and effects of a toxicant on reproduction, survival, feeding rate, or somatic growth rate. No test of results with real data.

1.3.3.4 Chaoborus IBM Population Model (Strauss et al. 2016)

Strauss et al. (2016) described an IBM for the phantom midge *Chaoborus crystallinus* to assess effect propagation of toxicants and subsequent population recovery. Individuals pass through four larval instars before they pupate and oviposit; adults outside the water phase are not explicitly simulated. To consider recolonization, the model can simulate two separate populations at the same time that are connected by the migration of adults (given proportions of adults are considered to oviposit locally, migrate to the other site, or are lost during migration).

If population size exceeds the (constant) food level in the model, individuals slow down their development, but do not starve (background mortality during life stage transitions and reproduction is not affected by environmental conditions). Instead, density-regulation is considered to be achieved through cannibalism, a unique feature in the Chaoborus IBM Population Model: The mortality of first instar larvae increases with the overall larval population density, and the susceptible time window of being first instar increases with food limitation due to the delayed larval development. However, preyed individuals do not contribute to the food supply of the survivors, so that the mechanism seems to rather to depict a crowding effect as implemented in IDamP (see section 1.3.3.3). It should be noted though that in IDamP, food limitation increases mortality and crowding delays individual development, whereas it is the other way around in the Chaoborus IBM Population Model. A combination of low water temperature and short photoperiod induces dormancy in winter in susceptible individuals; larvae do not develop and have a daily mortality risk while being dormant. The Chaoborus IBM Population Model can be linked to a dose-response or a GUTS module for the simulation of individual-level pesticide effects.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation Population recovery	
Environmental domain	Freshwater	Edge-of field ponds.
Taxon specificity	Specific	Phantom midge (*Chaoborus crystallinus*).
Toxicant specificity	Generic	
Application	Applied for registration	The model has been published recently but already proposed for risk assessment (personal communication).
Public availability	Scientific publication	No code or software publicly accessible.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Individuals Cannibalism	
Endpoints	Population size Population structure Survival	

Criteria	Categories	Comments
Space	Spatially implicit	Migration rate of adults between two simultaneously simulated populations can be set by the user.
Time	Days	Time step of 1 day. Usually run for months - years.
Exposure / effects	Varying concentrations Repeated exposure	Depends on the external exposure and toxicity module to which the population model is coupled.
Abiotic environment	Food limitation	The abiotic environment is represented by the food level (affects developmental rate), temperature (affects developmental rate and dormancy), and photoperiod (affects dormancy). Developmental rates of individuals have been calibrated to artificial ponds under Central European conditions.
Biotic environment	Intraspec. competition Cannibalism	Food limitation delays larval development, cannibalism increases mortality of first instar larvae.
Individuals	Stochastic	Sex, basic developmental rate, the initial developmental state, and the susceptibility to background mortality, cannibalism and dormancy are randomly assigned to each individual.
Populations	Individual-based	
Calibration	Laboratory data Mesocosm data	Extensive experimental work has been done to parameterize the effects of food level and population density on development and survival.
Programming language	Delphi® Professional 5.0.2.1.3	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Mesocosm data	Predicted population dynamics without pesticide exposure and after three successive exposure pulses to different concentrations have been validated in artificial ponds (when coupled with a GUTS module). Population recovery seems to be captured at low concentrations but overestimated at high concentrations.
Sensitivity analysis	Yes	
Uncertainty analysis	Yes	
Documentation	Scientific publication	Well-structured documentation following the ODD standard.

Assessment

Criteria	Description
Strengths	The model is simple and requires only few parameters, many of which have been calibrated using extensive experimental work in the laboratory and in artificial ponds. Predictions of the model for population dynamics without pesticide exposure and for population recovery after pesticide exposure have been validated with experimental data.

Criteria	Description
Theoretical uncertainties	The model is considered to simulate density regulation through cannibalism. However, the implemented mechanism (increase in first instar mortality with overall population density, but no benefit from cannibalism for the preying larvae) reminds rather on a crowding effect. Reproduction (number of offspring per adult) is considered independent from the environmental conditions experienced during larval development. Effects of interacting species are not considered.
Empirical uncertainties	A parameterization scheme for the food level has been experimentally developed that may need to be adjusted when other food sources in real ponds are considered.
Parametric uncertainties	Only few environmental factors modelled. No interaction of these stressors with pesticide effects at the individual level. Sublethal effects have not been incorporated yet.
Temporal uncertainties	Food supply is considered constant (in contrast to the seasonal cycles in water temperature and photoperiod).
Conclusions	The model appears useful to refine the risk of pesticides for the *Chaoborus crystallinus*. However, the model is quite specific for *Chaoborus crystallinus* which is neither a standard test species nor does it seem particularly vulnerable according to its ecological traits. Therefore, applicability in the European framework for ERA is limited. Additionally, the model does not explicitly simulate effects of interacting species that may delay population recovery. It seems that additional species were also not present in the studies used for calibration, so that they are also not covered implicitly. Accordingly, the model may underestimate the time for population recovery in the field.

Publication

Citations	Taxa	Chemicals	Comments
Model description			
Strauss et al. (2016)	*Chaoborus crystallinus*	None	Original publication of the model.
Model applications			
Dohmen et al. (2016)	*Chaoborus crystallinus*	Insecticide	Coupling with GUTs and validation of predicted population recovery from pulse exposure with mesocosm data.

1.3.3.5 IBM with DEB (Martin et al. 2012)

Martin et al. (2012) presented an example of an IBM coupled to a toxicity module that uses a energy budget for the calculation of individual-level effects and was applied to *Daphnia magna*. Use of the dynamic energy budget (DEB) theory in individual-level TKTD models has been described in section 1.2.2.2. DEB describes survival, growth and reproduction over the life cycle (and the effects of stressors on these traits).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation	
Environmental domain	Generic	
Taxon specificity	Generic	
Toxicant specificity	Generic	
Application	Established in science	Both DEB and the concept of IBMs are well-established in Science. Their combination is not too common.
Public availability	Software extension	Implementation in NetLogo.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Effects on individuals are treated by the DEB component, and they are propagated to population dynamics in the IBM.
Endpoints	Survival Reproduction Population size	Endpoints may include population numbers, biomass, structure or recovery.
Space	No spatial context	The published examples with DEB-IBM deal with a homogeneous environment, but an extension to a spatial setting is available on the website.
Time	Dates	
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure Toxicant mixtures	No limitations in terms of exposure scenario. Basic mixture effects have been added to DEB models and can also be included in DEB-IBM. Testing of DEB-IBM has so far been limited to constant exposure and single compounds.
Abiotic environment	Food limitation Temperature	Food is explicitly followed in DEB-IBM, which implies food limitation at high population densities. Temperature effects can easily be added.

Criteria	Categories	Comments
Biotic environment	Intraspecific competition	Only competition through food.
Individuals	Stochastic Energy-budget	Stochastic for survival, energy budget for growth and reproduction.
Populations	Individual-based	
Calibration	Laboratory data	DEB models are calibrated on individual-level data. Some model adaptations have been suggested to capture responses at low food in a population-level test.
Programming language	NetLogo	NetLogo is freeware. DEB-IBM is downloadable from https://popecology.wordpress.com/deb-ibm/

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	DEB-IBM was validated using population-level tests in the laboratory. DEB has been thoroughly tested itself.
Sensitivity analysis	No	
Uncertainty analysis	Yes	Several sources of uncertainty/variation are propagated in the model: mortality, assimilation rate, food supply.
Documentation	Scientific publication Website	Several publications are available, as well as a description using the ODD protocol (https://popecology.wordpress.com/deb-ibm/).

Assessment

Criteria	Description
Strengths	DEB-IBM is a combination of well-established theory for individuals, and well-established IBM calculation. As DEB provides a detailed description of an individual's life history, the connection to IBMs is natural. The DEB component is generic (not species or stressor specific) and allows the individuals to respond realistically to time-varying conditions (toxicants, food, etc.).
Theoretical uncertainties	Apart from the uncertainties in DEB, the DEB-IBM currently follows a population of one species in isolation (no interspecies competition, no predators, no parasites, etc.). This may not be representative of field populations.
Empirical uncertainties	Many species, or data sets, require adaptations to the DEB model; often difficult to find a unique mechanism of action of a chemical; no applications to birds and mammals. DEB-IBM for Daphnia includes a rather ad-hoc modification for starvation mortality.
Parametric uncertainties	Often difficult to identify all DEB parameters uniquely from standard data sets.
Temporal uncertainties	
Conclusions	DEB-IBM is a straightforward way to extend DEB modelling to population dynamics. The use of DEB makes this population model largely generic, and adds the possibility for individuals to respond realistically to changing conditions. The focus on a single population in isolation limits its realism for field situations.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Martin et al. (2012)	Generic	Generic	General presentation of the DEB-IBM model.
Model applications			
Martin et al. (2013a)	*Daphnia*	Food stress	Population-level test of the model.
Martin et al. (2013b)	*Daphnia*	Dichloroaniline	Population-level test of the model with toxicant stress.
Martin et al. (2014)	*Daphnia*	Various hypothetical compounds	Simulations for various mechanisms of action; general patterns.

1.3.3.6 IBM and PDE with kmDEB for Earthworms (Baveco and De Roos 1996)

Baveco and De Roos (1996) provided an IBM for earthworms coupled with a toxicity module for individual-level effects that is based on the Kooijman-Metz (km) DEB. kmDEB is a predecessor of DEB theory, and bares a close resemblance to DEBtox. It is based on a simplified energy budget. In comparison to DEBtox, kmDEB has no reserve, no maturity maintenance, and no TK considerations. Sublethal effects are included through a change in various energetic processes. For comparison, the kmDEB module was also coupled with a PDE (partial differential equation) based population model (see example in section 1.3.2.12), which does not consider a limited number of individuals and stochasticity. In the phase of population decline and initial recovery after exposure, both the IBM and PDE model yielded similar behaviour of the population size. In the IBM, however, fluctuations in population size remained in the equilibrium phase, mainly due to demographic stochasticity that acted on the density of adult individuals. In contrast, initial oscillations before the population models reach stable state disappeared faster in the IBM, due to the variability in cocoon incubation times leading to a faster spreading out of successive birth peaks.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	Authors placed the model in the context of PPP risk assessment.
Questions / processes	Individual effects Effect propagation Population recovery	Translation of effects in individuals to population dynamics, both under constant and pulsed exposure.
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	The individual-level model is generic (kmDEB), but the ecological context is quite specific for earthworms.
Toxicant specificity	Generic	Hypothetical compounds are used; no toxicokinetics, so instant steady state.
Application	Little-known	kmDEB is a predecessor of DEB theory, and closely resembles DEBtox. Here, it is linked to individual-based population methods (IBM and PDE).
Public availability	-	No implementations available.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Growth Reproduction	Sub-lethal effects on individuals are treated in the kmDEB module, and they are propagated to population dynamics with a PDE and IBM implementation.
Endpoints	Population size	Model calculates population size and structure over time as a result of constant or time-varying exposure. Also, recovery times and extinction probabilities are assessed.
Space	No spatial context	
Time	Dates	Life history of individuals is treated on a per-day basis, and population dynamics is followed over multiple years.

Criteria	Categories	Comments
Exposure / effects	Repeated exposure Chronic vs. pulse exposure Varying concentrations	Exposure profiles used are constant exposure, single application and repeated exposure. The model can deal with any exposure pattern.
Abiotic environment	-	Food is assumed to be available ad libitum and temperature constant. The kmDEB model has, in principle, the possibility to deal with both.
Biotic environment	Predation or intraspecific competition	Density-dependent predation, implemented as a mortality rate.
Individuals	Energy budget Stochastic	Energy budget for growth and reproduction, and mortality is stochastic in the IBM.
Populations	Individual-based	Both as PDE (infinite number of individual) and true IBM (finite number of individual and stochasticity). The two methods complement each other.
Calibration	Laboratory data	Literature data from various sources were used to calibrate the model.
Programming language	SmallTalk	The IBM was implemented in EcoTalk, which uses the SmallTalk framework.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation attempts. The kmDEB model was used in a number of other studies, and describes individual life history well.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	The model uses an individual module based on the dynamic energy-budget framework. Both a PDE and IBM implementation are used to assess the influence of stochasticity.
Theoretical uncertainties	The kmDEB model is not used anymore, as it has been replaced by DEBtox. This implementation does not consider toxicokinetics not food limitation. Predation is included as a density-dependent mortality.
Empirical uncertainties	
Parametric uncertainties	Parameterization of the energy-budget model is hampered by lack of data on earthworm life history under controlled conditions.
Temporal uncertainties	Environment (food, temperature, predation) is taken as constant, although this can be modified in the model.

Criteria	Description
Conclusions	The kmDEB model is outdated and has been replaced by DEBtox/DEBkiss. The PDE/IBM approach is powerful in showing effects of a single or multiple application of a toxicant, and the time needed for the population to recover.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Baveco and De Roos (1996)	Two species of earthworm	Hypothetical	
De Roos et al. (1992)	*Daphnia*	None	Simulations with the combination kmDEB and a PDE, solved with the Escalator-Boxcar Train approach.
		Model applications	
Rinke and Vijverberg (2005)	*Daphnia*	Environmental conditions	Model is very similar to kmDEB but takes assimilation efficiency and maintenance rate as function of food density. Used with the escalator-boxcar train method, so might be classified as a PDE.
Groeneveld et al. (2015)	Antarctic krill	Environmental conditions	DEBkiss (very similar to kmDEB) used as module for krill life history in an IBM population model.
Fiechter et al. (2015)	Chinook salmon	Environmental conditions	DEBkiss (very similar to kmDEB) used as module for salmon life history in an ecosystem model. Salmon population modelled with IBM.

1.3.3.7 IBM with NPM for *D. magna* (Vanoverbeke 2008)

The IBM of Vanoverbeke (2008) for *Daphnia magna* incorporated an energy-budget component for the individuals in the form of a net-production model (NPM; maintenance costs are paid from assimilation first, after which the remainder is used for growth and reproduction, see section 1.2.2.5). This differs from the net-assimilation approach in DEB theory, but is used in a large number of population models. A lot of modelling effort is placed on the response to food limitation, starvation and crowding, which turn out to be crucial for the population dynamics. Population dynamics is calculated using an IBM (following cohorts), with the algal food following logistic growth.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	The model is aimed to understand the population dynamics of Daphnia magna from the individual energetics and the relationship with its food.
Questions / processes	Others	The model is aimed to understand the population dynamics of *Daphnia magna* from the individual energetics and the relationship with its food. There is no toxicant stress included, although it might be added.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	The model is rather specific for *Daphnia magna*, although it might be re-parameterized for other filter-feeding zooplankters.
Toxicant specificity	-	No toxicants involved.
Application	Little-known	The model does not seem to be used beyond the initial publication, although it shares many similarities with other energy-based models (it is a net production model).
Public availability	-	The model does not seem to be available in public form.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	The model contains an energy budget for the individuals, and an IBM module for the population dynamics.
Endpoints	Survival Reproduction Population size	Both individual behaviour and population dynamics are outputs of the model.
Space	No spatial context	
Time	Days	
Exposure /	-	No toxicants included, although they could be added.

Criteria	Categories	Comments
effects		
Abiotic environment	-	
Biotic environment	Intraspec. competition	Algal food is treated as a simple population, growing according to a logistic growth function. Competition between individuals is including through the food source and though additional crowding effect.
Individuals	Energy budget	Individuals are described with an energy budget of the net-production type (maintenance costs are paid from assimilation first). All individuals have the same properties in the model. Several, rather descriptive, model elements are included to match observed behaviour in laboratory and field data. This makes the model rather parameter rich and specific for *Daphnia magna*.
Populations	Individual-based Logistic growth	*Daphnia* population dynamics is described by an IBM, following cohorts over time. The algal food is modelled with a simple logistic growth function.
Calibration	Laboratory data Field data	Calibration is done by taking parameter values from the literature, as well as calibrating a number of parameters to individual-level data (not clarified and not shown). Furthermore, several parameters/processes were adapted to match certain types of population behaviour (e.g., density dependence).
Programming language	C++	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Laboratory data	Some comparisons are made between individual traits and observed values (size, age, eggs at first brood), and qualitative comparison to observed population patterns. However, to some extent these actions should be regarded as calibrations.
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	The model is well documented, but the parameterization and calibration less so.

Assessment

Criteria	Description
Strengths	Detailed representation of *Daphnia* energetics, ability to reproduce certain types of population dynamics (low and high amplitude cycles).
Theoretical uncertainties	Algal food is treated as logistic growth, no interactions with other species (e.g., competition and predation).
Empirical uncertainties	-

Criteria	Description
Parametric uncertainties	The model includes a large number of parameters to capture individual life history of Daphnia.
Temporal uncertainties	-
Conclusions	A rather complex model that includes many details of *Daphnia* behaviour. As no fits/comparisons to data for growth, reproduction, ingestion, etc. are shown, it is difficult to judge the conceptual validity of the model at the individual level.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Vanoverbeke (2008)	*Daphnia magna*	None	Example of this model type that has been reviewed here.
Sinko and Streifer (1969)	*Daphnia magna*	None	One of the earliest NPMs. Linked to a partial-differential equation to extrapolate to population dynamics.
Model applications			
Peeters et al. (2010)	*Daphnia*	None	Somewhat different net-production model, calibrated to data for growth and reproduction at different food levels. Population simulations using Escalator-Boxcar Train.
Johnston et al. (2014b)	Earthworms	Copper and chlorpyrifos	Inclusion of the Sibley et al. NPM into an IBM and adding toxicant stress (this model has its own entry as population model).

1.3.4 Spatially Explicit IBMs

1.3.4.1 Spatial IBM for Marine Crustaceans (De Los Santos et al. 2015)

de los Santos et al. (2015) published a simple spatially explicit IBM to assess the propagation of lethal and sublethal individual-level effects of chemicals to populations of the marine crustacean *Gammarus locusta*. Toxicant effects on individuals are modelled with a simple dose-response module. In the model, individuals move randomly between grid cells; reproduction and growth depend on temperature and the exposure time and concentration. Mortality depends on local population density and toxicant concentration. Grid cells only differ in local population density and toxicant concentration. The model was applied to aniline that increased mortality and decreased reproduction in individual-level tests. Even low concentrations (0.5 – 2.5 µg/L) resulted in long-lasting (up to 200 – 500 days) population recovery times when the exposure period exceeded 20 days.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation Population recovery	Assess individual effects of chronic exposure to aniline and their propagation to population level; estimate recovery time.
Environmental domain	Marine	
Taxon specificity	Taxon-specific	*Gammarus locusta*
Toxicant specificity	Generic	Applied to aniline
Application	Little-known	
Public availability	Source code	

Variables and Parameters

Criteria	Categories	Comments
Entities	Individuals	Female individuals and grid cells.
Endpoints	Population size Recovery time	
Space	Grid cell	In the application 25 cells of 1 m².
Time	Days	Application proceeds in daily time steps and was run for 4 years.
Exposure / effects	Homogeneous Dose-response	Homogeneous exposure in all cells. Individual-level effects from dose-response models fitted with acute and chronic standard test data for the simulated exposure times.
Abiotic environment	Temperature	Temperature homogeneous in all cells, updated daily based on external weather data; affects growth and reproduction.
Biotic environment	Intrasp. competition	Density-dependent mortality calculated for each cell.

Criteria	Categories	Comments
Individuals	Stochastic	Only females modelled: Age, body length (juveniles or adults based on size), brood size, embryo age; initial properties of each individual drawn from probability function.
Populations	Individual-based	
Calibration	Field data Laboratory data	Initial population density and structure based on field observations. Temperature regime linearly interpolated from monthly average temperatures. Mortality based on laboratory studies. Effect of aniline from acute and chronic toxicity tests.
Programming language	R 3.0.2	R package simecol.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Only validation of population dynamics without exposure. Predictions from 5 simulations overestimated population density but met population dynamics from field data.
Sensitivity analysis	Yes	Each parameter varied by +/- 10 %, 5 simulations per parameter combination. Parameters that affected the density-dependent mortality, reproduction and growth had the highest effect on mean annual population density.
Uncertainty analysis	No	
Documentation	Scientific publication	Scientific publication and description according to the ODD protocol.

Assessment

Criteria	Description
Strengths	Low complexity, few parameters to be estimated.
Theoretical uncertainties	No species interactions, no food limitation, no abiotic stressors, no sublethal or chronic (delayed) effects, no migration.
Empirical uncertainties	Sensitivity of different size or age classes not considered.
Parametric uncertainties	Use of grid cell ungrounded in the application as no spatial heterogeneity of the environment was modelled, and Gammarus tends to distribute homogeneously across the modelled 25 m² to minimize density-dependent mortality. Number of offspring depends only on the size of the mother.
Temporal uncertainties	Interpolation of mean monthly temperature ignores extreme temperatures that can have important effects.
Conclusions	Very simple model with few parameters. Few environmental factors considered that may interact with the effects of toxicants on the population growth, therefore time of recovery is likely underestimated. This has not been tested.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
de los Santos et al. (2015)	*Gammarus locusta*	Aniline	Original description of the model.

1.3.4.2 **MASTEP (Van den Brink et al. 2007)**

MASTEP (Van den Brink et al. 2007a) is another example of a simple spatially explicit IBM. It has been developed to assess population recovery of *A. aquaticus* through reproduction and migration after pulse exposure to pesticides in agricultural freshwater bodies. Main motivation was to increase realism of mesocosm studies by virtually repeating them with the addition of immigration from non-exposed stream stretches. In the model, pesticides eliminate a given fraction of the individuals in exposed cells of a quadratic lattice using a built-in dose-response module or alternatively an external GUTS module. The model provides a built-in fate module that can simulate pesticide drift under consideration of water flow velocity.

Individuals walk randomly among landscape cells and may occasionally drift downstream, facilitating recovery in a previously exposed area. Landscape cells differ only in local population density, pesticide concentration, and accessibility type (individuals can enter water but not land cells). The modelled life cycle is very simple: Times of reproduction and death due to old age are randomly scheduled at birth. Density stress (other individuals present in the same landscape cell) linearly increases the daily random background mortality. The history of experienced density stress (average number of individuals present at the same cell since birth) decreases brood size. Individuals try to avoid density stress, as presence of other individuals increase the probability of movement to a neighboring cell (decrease in the random residence time). Parameter estimates were based on expert judgment and on a review of published information on the ecology of *A. aquaticus*.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	Risk assessment of chemicals.
Questions / processes	Population recovery	Assessment of population recovery after pesticide exposure and its dependence on landscape configuration.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	*Asellus aquaticus*; was also parameterized for *Gammarus pulex* in later applications.
Toxicant specificity	Generic	Only for acute effects.
Application	Known in science	
Public availability	Source code Stand-alone program	Demo for *A. aquaticus* is available at https://www.mastep.wur.nl/documentation_demo.shtml

Variables and Parameters

Criteria	Categories	Comments
Entities	Individuals	
Endpoints	Population size Recovery time	Development of population size after pulse exposure.

Criteria	Categories	Comments
Space	Grid cell	1 cell = 1 m². The last cells of the grid are connected to the first cells (periodic boundary conditions). Scenarios have been shown with a ditch of 600 m length, and also extended model versions with several km² landscapes.
Time	Days	Daily time steps. Simulations were run for at least 1 year.
Exposure / effects	Varying concentrations	MASTEP has been connected to TOXSWA and FOCUS exposure models for pesticide loadings. The model includes an internal fate module for drifting within the modelled landscape and degradation in each landscape cell. Authors recommended to parameterize acute effects (dose-response submodel for mortality) in the model with observed mortality in a mesocosm study ca. 7 days after exposure. This way, short-term delayed effects and effect interactions with additional stressors at individual level are implicitly covered, but the model is limited to the same exposure profiles as those in the study used for parameterization. Later, MASTEP was also coupled to a GUTS module (Dohmen et al. 2016).
Abiotic environment	-	No abiotic conditions except that landscape cells are assigned as water / no water (only water can be populated).
Biotic environment	Intraspec. competition	Currently experienced density stress increases background mortality and the probability of moving to another cell. History of experienced density stress decreases litter size.
Individuals	Stochastic	Time of reproduction and natural death, sensitivity to density stress and basic number of offspring are drawn for each individual from probability functions.
Populations	Individual-based	
Calibration	-	No calibration, parameters were taken from literature (breeding, background mortality and aging) or based on expert judgement (movement and sensitivity to density stress).
Programming language	Netlogo 4.1	Model was originally developed in VisualWorks Smalltalk using the EcoTalk modeling framework. The available online version is implemented in NetLogo.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation.
Sensitivity analysis	Yes	Drift parameter is important for the recovery of contaminated section, but not for recovery of the whole stream.
Uncertainty analysis	Yes	95 % confidence intervals from probabilistic model runs are reported.
Documentation	Scientific publication Website	Well-structured and comprehensive documentation following ODD protocol. Website with demo: https://www.mastep.wur.nl/documentation_demo.shtml

Assessment

Criteria	Description
Strengths	The simple modelled life cycle requires only few input parameters. Population dynamics emerge from simple and logical rules and considers variation between individuals.
Theoretical uncertainties	No sublethal or chronic effects and no species interactions modelled, therefore population recovery from reproduction may be overestimated. Acute effects may not be captured well for scenarios that differ from the one used for parameterization of individual-level effects (e. g. higher density stress or harsher abiotic conditions may increase sensitivity to pesticides).
Empirical uncertainties	High uncertainty regarding dispersal parameters and the density-dependence of life history traits, which are difficult to measure.
Parametric uncertainties	No discrimination between life stages (young individuals may be more sensitive than older ones, and more demographic effects may arise when different life stages compete for different resources). Environment is modelled homogeneous, though spatial heterogeneity (e. g. in food availability) may increase aggregation and thus local density stress in real populations.
Temporal uncertainties	Environment is assumed to be constant. If the dose-response module is used for individual-level effects, acute effects are valid only for the exposure profile used in parameterization, though in the simulations drift may result in different exposure profiles across the landscape cells.
Conclusions	MASTEP may support mesocosm studies when the experiment is virtually repeated with additional immigration from non-exposed stream stretches. Simulations of other scenarios should be interpreted with care because many processes that potentially affect the acute effects and the population recovery are not explicitly modelled.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Van den Brink et al. (2007a)	*Asellus aquaticus*	Generic	First publication.
Van den Brink and Baveco (2009)	*Asellus aquaticus*	Generic	Concise summary in a textbook.
http://www.mastep.wur.nl/			Online documentation and demo version.
Model applications			
Galic et al. (2012)	*Asellus aquaticus*	Generic	Analysis of effects of landscape composition on population recovery.

Citations	Taxa	Chemicals	Comments
Baveco et al. (2014)	*Asellus aquaticus* *Gammarus sp.* *Chironomus sp.* Ephemeroptera	Generic	Comparison of outcome when MASTEP has been parameterized for 4 species with different life histories and coupled to different modules for individual-level effects (dose-response and TDM (equivalent to GUTS-IT)). Also comparison with outcome of non-spatial model. Recovery took longer in the spatial models than in the non-spatial model when there was spatially heterogeneous exposure and little movement; otherwise no relevant differences.
Focks et al. (2014a)	*Asellus aquaticus*	Generic	Analysis of recovery times when concurrent or sequential exposure to multiple pesticides is simulated.
Focks et al. (2014b)	*Asellus aquaticus*	Generic	Linking MASTEP to the fate models CASCADE-TOXWA and extension to a regional approach (10 km²).
Galic et al. (2014)	*Gammarus pulex*	Anonymous insecticides	Re-parameterization to *Gammarus pulex* and coupling with TDM module for individual-level effects (can predict delayed mortality). Comparison with performance when a dose-response module is used.
Dohmen et al. (2016)	*Gammarus pulex*	Anonymous insecticide	MASTEP and two other IBMs (Chaoborus IBM, IDamP) were linked to GUTS module, and a case study was simulated. Comparison of model outcome with conventional risk assessment and a mesocosm study.

1.3.4.3 IBM with GUTS for Aquatic Invertebrates (Baveco et al. 2014)

In the model of Baveco et al. (2014), an IBM approach has been combined with a TKTD model for the effect on the individual's survival (TDM, which is a special case of GUTS) to propagate effects to the population level. This combination was first presented by Ashauer (2010). However, in that paper, the IBM is extremely basic (just a large number of individuals with different parameter values that are followed over time in a constant homogeneous environment). In Baveco et al. (2014), it is more elaborate including density dependence, spatial context, and more detailed description of the life cycle (incl. reproductive events and dispersion). Baveco et al. (2014) compared predictions of this model (IBM-TD) for four invertebrate species with different life history traits with results from an IBM coupled to a simple dose-response module (IBM-DR) and with results from a non-spatial logistic growth model coupled to a dose-response module (Log-DR). Recovery from spatiotemporally homogeneous pulse exposure was quite similar in the IBM-DR and Log-DR, only for the mayfly (characterized by low mobility) the Log-DR predicted faster recovery because reproduction in the spatial IBM was hindered by high local population density following clustered reproduction. After spatiotemporally heterogeneous exposure, recovery times in the IBM-DR were longer than in the Log-DR coincided for species that were not highly mobile. Recovery times in the IBM-TDM were shorter or longer than in the IBM-DR depending on the exposure scenario and the species traits.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation Population recovery	A TKTD model for the effect on the individual's survival (GUTS-based) is combined with an IBM approach to propagate effects to the population level.
Environmental domain	Freshwater	The TKTD module is generic, but the IBM context is more specific for the species modelled (four species of aquatic invertebrates).
Taxon specificity	Taxon-specific	
Toxicant specificity	Generic	The models could be used for other substances, but only if mortality (which is the only trait considered) is the dominant effect.
Application	Little-known	The GUTS model used as individual-level module is well known, but the combination with IBM does not have a large distribution so far.
Public availability	Software extension	The models do not seem to be publicly available.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Mortality Reproduction	The IBM model includes reproduction and survival for individuals, and dispersion to other grid cells. Only effects on survival are included.

Criteria	Categories	Comments
	Population density	
Endpoints	Population size Recovery time	The model is able to calculate population densities and recovery times.
Space	Grid cell	Model landscape is made up of 1m-squared grid cells, with 600-900 cells in total.
Time	Days	Basic time step in the IBM was 1 day, and the population was followed over several years.
Exposure / effects	Chronic vs. pulse Varying concentration Repeated exposure	Exposure profiles were generated with TOXSWA, for different FOCUS scenarios.
Abiotic environment	-	The environment is assumed to be constant: no changes in temperature or food level are included in the model. Another version within the GUTS-IBM category (Diepens et al) does consider seasonality.
Biotic environment	Intraspec. competition	Density dependence was included via the mortality rate.
Individuals	Stochastic	The individuals follow a pre-programmed life history over time, with stochasticity on the timing-related parameters (age at reproduction and time between broods). Rules for mobility were included, depending on the species modelled. The only effects of the chemical are through the survival probability.
Populations	Individual-based	Individuals are followed in the landscape.
Calibration	Laboratory data Field data	Data for 4 freshwater invertebrates from laboratory were used to parameterize the life history and the GUTS module for survival effects. Field observation seem to be used to estimate some parameters regarding the life cycle.
Programming language	NetLogo	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	A description according to the ODD protocol is available as supplementary information.

Assessment

Criteria	Description
Strengths	The TKTD module allows individual effects to respond realistically to the time-varying exposure concentration. The IBM contains species-specific details on movement of the individuals.

Criteria	Description
Theoretical uncertainties	Only effects on survival are included, whereas sub-lethal effects will generally appear at lower concentration levels. No other species are included (predators, competitors, parasites, etc.). Further, the uncertainties of the GUTS model apply here as well.
Empirical uncertainties	It will be difficult to represent complex life histories (and their dependence on the environment) and mobility into simple parameters (such as rate constants).
Parametric uncertainties	TK and TD parameters for individuals were kept constant, i.e., no effect of body size/growth on TK. The model of Diepens et al. does consider growth but assumes exponential growth. No effects on temperature on the rate constants in the TKTD model. The implementation by Baveco et al. includes no effects of temperature or seasonality on the life history.
Temporal uncertainties	Environment is taken as constant over the year (the model of Diepens et al. (2016) does consider seasonality, though).
Conclusions	The combination of GUTS with IBMs is useful to provide insights into the population-level impacts of effects on individual survival probability. The model of Diepens et al. increases realism by including ingestion in the TK model, and seasonality. More realism may be added by including size-dependent TK. However, these models will remain limited to effects on survival.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Ashauer (2010)	*Gammarus*	Chlorpyrifos and pentachlorophenol	First publication on the link between GUTS and an IBM, although the IBM component is very simple (mainly intended to show the influence of inter-individual differences in parameters on survival probability over time).
Baveco et al. (2014)	4 species of aquatic invertebrates	Chlorpyrifos	Main publication reviewed here.
		Model applications	
Diepens et al. (2016)	*Chironomus*	Chlorpyrifos	A more extended IBM model, including seasonality and an extended TKTD module (including uptake from ingestion and dilution by growth).
Galic et al. (2014)	*Gammarus*	4 pesticides	The IBM includes the effect of temperature on life history, with a realistic temperature profile over the year.
Gabsi et al. (2014b)	*Daphnia*	Dispersogen A	Link of GUTS to the iDamP IBM.
Dohmen et al. (2016)	*Gammarus, Chaoborus, Daphnia*	Hypothetical	Linking the GUTS model to three different (previously published) IBMs, and case study in context of ERA for PPPs.

1.3.4.4 **SpringSim (Meli et al. 2013)**

Meli et al. (2013) provided a comparably spatially explicit IBM for springtails that is more complex as compared to the spatially explicit IMBs for aquatic organisms described above. Individuals are characterized by their age, position, direction for movement, energetic status (time-to-death without food intake), cumulative distance walked in each time step (affects energy used for movement), and time spent on contaminated grid cells. The individuals actively search for and consume food items in a heterogeneous environment and avoid highly contaminated cells, resulting in an aggregation in refuge areas. The amount of individual movement varies with the seasonally changing temperature. Contamination increases mortality based on a simple log-linear dose-response module. Due to refuges, the model predicts a higher overall population size at the same mean concentration if the contamination is spatially heterogeneous.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	Risk assessment of chemicals.
Questions / processes	Spatial heterogeneity	How does spatially heterogeneous contamination affect toxicity effects?
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	*Folsomia candida*; developed for springtails but may be adapted to other species.
Toxicant specificity	Generic	Applied to copper sulfate (CuSO4), but applicable to other substances.
Application	Little-known	
Public availability	Source code	

Variables and Parameters

Criteria	Categories	Comments
Entities	Individuals	Eggs, juveniles, female adults (stage is age-dependent).
Endpoints	Survival Population size	
Space	Grid cell	100 * 100 cells, each representing 1 cm².
Time	Hours - days	6 seasons corresponding to different temperature ranges. The order in which the individuals are processed is randomized each time step [h]. Some processes are performed only each day.
Exposure / effects	Varying concentrations	Survival depends on the toxicant concentration and the amount of time the individual spends on contaminated patches. Concentration-dependent reduction of egg hatching success.
Abiotic environment	Food limitation Temperature	Food is heterogeneously distributed in random "food cells" and is restored every day.

Criteria	Categories	Comments
Biotic environment	Intraspec. competition Crowding	Reduction of fecundity through high population density (crowding) on a grid cell. Food limitation decreases survival and reproduction.
Individuals	Stochastic	Springtails characterized by their life stage (egg, juvenile, female adult, depending on age), position, direction for movement, energetic status (time-to-death), and cumulative distance walked in each time step (affects energy used for movement). Trait values of each individual are randomly drawn from distributions at birth. Each season individuals get a different set of parameter values reflecting temperature-dependent development. Adults reproduce according to the parameters "time between broods" and "number of broods". Individuals sense food availability and concentration of the local grid cell and the cell ahead and prefer non-contaminated, food-rich cells.
Populations	Individual-based	
Calibration	Laboratory data	Pattern-oriented calibration of energy-related parameters, i. e. search of parameter values that reproduced two patterns of population dynamics observed in microcosm studies (food-dependent and density dependent population growth). The other parameters were taken from studies in the literature.
Programming language	Netlogo 5.0	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Mesocosm data	The calibrated model was tested with independent patterns of population dynamics (generation time, seasonal variation in population size, intrinsic population growth rate r) from greenhouse experiments. Good fit of r, relatively good fit of seasonal population size.
Sensitivity analysis	Yes	Sensitivity analysis for parameters that had to be calibrated from the model by changing the values by ± 10 to ± 50 %. Season parameters were held to a constant to the spring / fall values.
Uncertainty analysis	-	
Documentation	Scientific publication	Comprehensible description according to the ODD standard.

Assessment

Criteria	Description
Strengths	Realistic implementation of processes within the springtail population, such as contaminant avoidance and foraging instead of random walk. Most parameters taken from literature, thorough calibration of the remaining energy-related parameters. Pattern-oriented modelling including validation with independent data demonstrates structural realism of the model.

Criteria	Description
Theoretical uncertainties	No interspecific interactions, no delayed (chronic) effects of a toxicant, or indirect effects through the food web. Predation and predator-avoiding behaviour not considered, but may force individuals also to aggregate in contaminated areas to hide from predators.
Empirical uncertainties	Energy-related parameters had to be calibrated from the model.
Parametric uncertainties	Only a very coarse energy budget represented (days of survival until starvation), therefore only crude estimation of sublethal toxicant effects (reduction of fecundity).
Temporal uncertainties	No migration. Under repeated exposure, models without interspecific interactions tend to produce equilibrium population sizes, while the pressure from antagonistic species may drive the weekend population finally to extinction (culmination).
Conclusions	A realistic model of isolated populations under experimental conditions. Useful to compare effects of different exposure scenarios on springtail populations, but low transferability of predictions to the field due to missing species interactions.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Meli et al. (2013)	*Folsomia candida*	Generic	Original publication.
Model applications			
Meli et al. (2014a)	*Folsomia candida*	Generic	Virtual experiments to assess the combined effects of habitat fragmentation, pesticide exposure and natural stressors.
Meli et al. (2014b)	*Folsomia candida*	Generic	Comparison of the model with a matrix model (RAMAS). Similar predictions when exposure is spatially homogeneous, more precise predictions of Meli et al. under heterogeneous conditions.
Reed et al. (2016)	*Folsomia candida*	Generic	Application to a hypothetical exposure scenario to discuss the potential use in risk refinement (together with Johnston et al. 2014b); weather data from FOCUS included.

1.3.4.5 Spatial IBM with Energy Budget for Earthworms (Johnston et al. 2014)

Johnston et al. (2014b) provided a comparably complex spatially explicit IMB for the earth worm *Aporrectodea caliginosa* to mechanistically link effects on individuals to responses in field population. Landscape cells differ by soil water potential and the by the availability of food that is consumed and replenished. Individuals sense food and soil water potential in neighbouring cells and actively move to cells with best conditions. Low soil water potential affects ingestion rates and determines the onset of a resting phase (aestivation). Each individual has its own energy budget that follows principles of physiological ecology and is based on a different energy-budget philosophy than DEB: Under suboptimal conditions, reproduction is prioritized over growth. The global soil temperature varies seasonally and affects various process rates. Pesticides can affect survival, growth and reproduction through changes in the energy budget. Predictions of the energy budget part on the growth of body mass and cocoon production have been successfully validated with laboratory data under variable environmental conditions. Predictions of the population part on the distribution and abundance of individuals have been validated in spatially heterogeneous soil profiles without toxicant exposure.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Individual effects Effect propagation Population recovery	
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	So far, this model has only been applied to earthworms in terrestrial systems. However, the underlying energy budget is claimed to be generic.
Toxicant specificity	Generic	
Application	Little-known	The energy-budget module was presented in 2013, and the first application in an IBM context in 2014.
Public availability	Software extension	Implementation in NetLogo.

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth Reproduction Population density	Effects on individuals are treated by the energy-budget component, and propagated to population dynamics in the IBM.
Endpoints	Survival Reproduction Population size	
Space	Grid cell	2D lattice of quadratic grid cell, each covering 0.01 m².
Time	Days	The model proceeds in daily time steps.

Criteria	Categories	Comments
Exposure / effects	Chronic vs. pulse exposure Varying concentrations Repeated exposure	No limitations in terms of exposure scenario. However, no toxicokinetics are not included in the model (effect on a trait is instant and constant).
Abiotic environment	Food limitation Temperature	Food is explicitly followed in the energy budget; temperature affects the various rates through the Arrhenius function. Grid cells are characterized by food availability, food quality, soil temperature, soil water content and soil texture (both used to calculate soil water potential).
Biotic environment	Intraspec. competition	Only competition through food. Individuals need a mate to be able to reproduce.
Individuals	Energy budget	All individuals have their own energy budget (which follows a different philosophy from DEB theory). Individuals are characterized by life cycle stage (cocoon, juvenile or adult), mass and energy reserves.
Populations	Individual-based	
Calibration	Laboratory data	Parameters for earthworms were obtained from the literature. Some parameters were fitted to lab data (such as the von Bertalanffy rate constant).
Programming language	NetLogo	NetLogo is freeware.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data Field data	The individual model was compared to some lab data for growth and reproduction in earthworms. Predicted population density/biomass was compared to observed values in a field test.
Sensitivity analysis	Yes	Local sensitivity analysis of model parameters on population predictions (adult/juvenile biomass) and individual traits (reproduction).
Uncertainty analysis	No	No propagation of uncertainties.
Documentation	Scientific publication	Several publications are available, as well as a description using the ODD/TRACE protocol.

Assessment

Criteria	Description
Strengths	Explicit energy budget for the individuals. In principle, this allows for a mass balance (ingested food is removed from the environment and used to fuel growth and reproduction of the individuals). With the data sets in the paper, the model provides a good description of individual behaviour and results in realistic population dynamics.

Criteria	Description
Theoretical uncertainties	The individual-level model is not well tested under controlled conditions. It is therefore unclear whether it can provide a realistic behaviour of the individuals. There are several inconsistencies in the model, e.g., the reserves do not contribute to body mass, and overhead costs of transformations does no contribute to respiration. The IBM follows a population of one species in isolation (no interspecies competition, no predators, no parasites, etc.). This may not be representative of field populations.
Empirical uncertainties	The energy-budget model requires a number of parameters to be specified, which are difficult to establish (e.g., energy content of tissue and food).
Parametric uncertainties	May be difficult to identify all energy-budget parameters uniquely from the literature/experimental testing. The individual model lacks a TK component, which limits its application to compounds with fast kinetics in the organism.
Temporal uncertainties	
Conclusions	The energy-budget component has not yet been proven to provide a realistic representation of individuals under relevant conditions. The focus on a single population in isolation limits its realism for field situations. At this moment, work is being conducted in Aachen to compare the energy-budget component of the model to a DEB model and to measured data to obtain a better indication of each model's performance.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Johnston et al. (2014b)	Earthworms	Copper oxychloride and chlorpyrifos	First complete version of the model and inclusion into an IBM context.
Sibly et al. (2013)	Generic	None	Basic model for the energy budget of individuals.
		Model applications	
Johnston et al. (2015)	Earthworms	None	Effects of environment and management practices for populations.
Van der Vaart et al. (2016)	Generic	None	General discussion and possibilities for calibration.
Reed et al. (2016)	Earthworms	Hypothetical	Case study for applying the model in an ERA setting for a plant protection product.

1.3.4.6 **eVole (Wang et al. 2013)**

This spatially explicit IBM was first developed by Wang and Grimm (2007) for the common shrew and later adapted and extended by Wang (2013) for the common vole. The relatively simple model simulates population dynamics of territorial individuals in a landscape of hexagonal cells characterized by local food value, shelter (sufficient vegetation height), and pesticide concentration in the food. In contrast to the previously reviewed models, density regulation does not act through an increase in mortality with food shortage or high population density, but emerges from a set of behavioural rules: Adult females can only reproduce in the breeding season when they have established a home range that consists of a number of connected landscape cells with a sufficient amount of food and overlaps with a male's home range. Individuals compete for home ranges: They preferentially add neighbouring cells to their home range that are not part of another home range yet, and adult females expel each other from landscape cells. When the home range does not provide enough food (only cells with sufficient shelter can be used), individuals do not starve but start to wander randomly in search of vacant landscape cells to establish a new home range. The authors assumed that the home range-based approach adds realism and leads to more accurate predictions of population dynamics, especially in the contexts of pesticide risk assessment. Pesticide exposure of an individual is averaged across the food concentrations in the cells of it home range. Individual-level effects need to be calculated in an external module (built-in modules were implemented in later versions of the model).

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Effect propagation Population recovery	The first publication based on this model Wang and Grimm (2007) focused on the common shrew (*Sorex araneus*) and aimed to establish a model that describes home range dynamics of this species. A second publication by Wang (2013) aimed at validating that this model (with small modifications) is suitable to describe population dynamics of the common vole (*Microtus arvalis*). Bastiansen et al. (2013) presented a poster on an application of this model for risk assessment of sulfoxaflor. Finally, Schmitt et al. (2015) used this model to assess how a hypothetical fungicide affects populations of common voles.
Environmental domain	Terrestrial	The model has been parameterized for the common shrew and the common vole. Parameterizations for other species could be possible if appropriate data is available.
Taxon specificity	Taxon-specific	Wood mouse, common vole.
Toxicant specificity	Generic	The model requires input from Tier 1 risk assessment studies.
Application	Applied in studies	Four known scientific application, one includes risk assessment of a specific substance (Bastiansen et al. 2013); 80 citations of Wang and Grimm (2007) in google scholar (by June 2016).
Public availability	-	Not publicly available; maintained by RIFCON GmbH.

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density Reproduction Mortality	In addition to vital rates the model requires as input (1) a landscape with foraging values of habitat types, and (2) parameters that determines how behavioural rules generate home range dynamics.
Endpoints	Population size Reproduction Recovery time Spatial distribution	The model does not provide a fixed number of endpoints and is instead flexible in producing any endpoints related to the simulated spatio-temporal dynamics. The two risk assessment applications used population densities as endpoints.
Space	Grid cell	Hexagonal cells, each 16.24 m² (5m diagonal).
Time	Days	Daily time steps, total duration can differ. The risk assessment application was run for 1 year and 30 years.
Exposure / effects	-	The population model needs to be coupled to external models that provide toxicant exposure in time and space and the corresponding effects at the individual level. Since development of individuals (except of aging) is not explicitly modelled, implementation of developmental effects is difficult.
Abiotic environment	Food limitation Shelter	Landscape cells differ in food level and a sufficient food supply in a home range is required for reproduction. Only cells with sufficient vegetation cover (shelter) can be used for home ranges.
Biotic environment	Intraspec. competition	Individuals compete for home ranges, which indirectly drives reproduction.
Individuals	Stochastic	Individuals are characterized by sex and age that determines their life stages (pups, juveniles, adults). Processes such as death, reproduction, dispersal and home range establishment are stochastic or follow rules with stochastic components.
Populations	Individual-based	Population dynamics are an emergent property of the processes happening at the individual level.
Calibration	Field data Laboratory data	Parameterization was based on data on life history, survival rates, food availability for different habitat types, home range sizes and maximum dispersal distance.
Programming language	Unknown	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Predicted population dynamics of the model without pesticide applications have been partly validated. The model was able to reproduce observed reproduction, survival, spatial behaviour and population cycles of common voles. No validation of predictions on contaminated populations.
Sensitivity analysis	Yes	
Uncertainty analysis	No	No probabilistic modelling results have been shown, but may be created.

Criteria	Categories	Comments
Documentation	Scientific publication Website	The documentation is well structured, it follows the ODD protocol.

Assessment

Criteria	Description
Strengths	Individual-based model that explicitly considers home range dynamics and the related impacts on population dynamics. Simple model concept.
Theoretical uncertainties	Several ad hoc assumptions have been made that reflect limited understanding of the system, see parametric uncertainties for details.
Empirical uncertainties	Field data for calibration and validation of survival rates and spatial behaviour were limited.
Parametric uncertainties	Toxicant exposure and effects are not part of the model. Several behavioural rules related to home range dynamics and dispersal are ad-hoc assumption for which no observational data is available (e.g. dispersal as random walk or the details of the optimization algorithm underlying home range dynamics). Body weight is not considered and therefore toxicological effects on body weight cannot be incorporated. Potential inconsistency in assumptions regarding home range dynamics and survival rates: Individuals can have home ranges with suboptimal food amount (down to 40% of saturation threshold for home range establishment) but such a reduction in resources does not impact survival or reproduction.
Temporal uncertainties	No weather effects that could lead to differences in population development between years.
Conclusions	An individual based model that explicitly describes home range dynamics of individuals. The simple model design facilitates communication and understanding of modelling results. However, it is unclear to which extent the incorporation of home ranges is useful for the purpose of toxicant risk assessment. Home range and population dynamics for common voles seem to be captured fairly well, but the underlying behavioural rules are still based on several ad hoc assumptions.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Wang and Grimm (2007)	Common shrew	None	Presentation of the basic model for the common shrew.
Wang (2013)	Common vole	None	Adaptation of the model to the common vole and validation of population dynamics without toxicant exposure.
Model applications			
Bastiansen et al. (2013)	Common vole	Sulfoxaflor	Implementation of small effects on fertility leads to minimal impact at population level.
Schmitt et al. (2015)	Common vole	Hypothetical fungicide	Comparison of modelling results with outcome of classical risk assessment procedure to demonstrate suitability of IBMs for risk assessment.

1.3.4.7 **IBM with TK for the Wood Mouse (Liu et al. 2013)**

Liu et al. (2013) developed an individual-based model for the wood mouse as focal species for the risk assessment of pesticides. The model shares many features with eVole (see section 1.3.4.6) but is more complex. Quadratic landscape cells in the model are characterized by pesticide concentration, the current farming activity (affects survival), habitat quality (determined by plant cover and height), and the existence of a burrow system. Each cell can be part of one of 7 habitat types that determine the seasonal variation in plant cover and height and farming activity; the habitat type of a cell can change across years to simulate crop rotation.

In contrast to eVole, not only home ranges, but also the position of nests and the movement pattern of foraging individuals within their home range are explicitly modelled. Home ranges are not adapted continuously, but every 10 – 30 days, all individuals randomly acquire a new home range of a fixed size around their nest that is considered to remain constant during the following period. Home ranges can overlap. Individuals can sense habitat quality and local population density of cells within their home range. Every day, they forage at preferred sites with sufficient vegetation cover and low local population density within their home range, and move their nest to a better covered site if available. Individuals also occasionally visit random sites outside the home range, whose quality they cannot sense in advance. The daily survival decreases with increasing movement distance and local and global population density, with farming practices, and with decreasing plant cover of nests. Adults can reproduce during the breeding season if they have a nest site with high cover.

All the sites an individual has visited are recorded to calculate the exposure history, i. e. the proportion of foraging time spent in treated sites (PT) and the overall amount of ingested pesticide (assuming a given uptake rate while foraging). To consider also the temporal pattern of feeding, Liu et al. (2014) added an optional 1-compartment TK module that calculates the body burden of each individual based not only on ingestion, but also on absorption and elimination. Effects are related to the body burden using a simple dose-response relationship. Liu et al. (2014) compared the exposure predicted by the model with and without spatial heterogeneity in pesticide exposure, and with or without the TK module that introduces temporal variation in exposure due to varying patterns of ingestion). Spatial or temporal heterogeneity reduced the risk quotient (exposure divided by LD_{50}) by 37 – 85 %; the combination of both sources of heterogeneity had little further effect.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	Modelling aim was to realistically address potential risk of agricultural pesticides to mammals. The main purpose for building a TK-IBM was to understand better how the TK processes interact with spatio-temporal patterns of foraging, i.e. the main route of exposure.
Questions / processes	Body burden Effect propagation Population recovery	IBM in which internal pesticide concentrations can be calculated using toxicokinetic (TK) models, that are quantitative representations of the amount ingested as well as absorption, distribution, metabolism, and excretion (ADME).
Environmental domain	Terrestrial	Wood mouse (*Apodemus sylvaticus*)

Criteria	Categories	Comments
Taxon specificity	Taxon-specific	
Toxicant specificity	Generic	3 hypothetical toxicants simulated. All three pesticides had the same LD50 and were identical except that the elimination half-life ($t1/2= \ln 2/k_e$) was 30 min, 60 min or 120 min; consequently, the LC50s also varied.
Application	Known in science	WoS 4 citations (by June 2016).
Public availability	Source code	May be available upon request.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues Mortality Growth Reproduction	Mice age/develop and weaned juveniles leave the parental nests. Adults and juveniles acquire (potential) home range and forage. In addition to plant cover, energy requirements and probability of eating newly drilled seeds are taken into account in foraging. Pesticide exposure is calculated and compared with lethal thresholds (individual tolerance); survival is affected by the pesticide-induced mortality. Mice have a certain probability to go excursion and then go back home and change nest if necessary. Reproduction occurs in nest site. Survival also depends on the daily background mortality.
Endpoints	Population size Recovery time	Population-level effects, i.e. mean dose and internal toxicant concentration, population size and mortality, are also outputs from the model. Time to recovery was calculated, but data not shown.
Space	Grid cell	The total size of the landscape is 10.4 ha, which is represented as 101 × 41 square patches, with torus setting to avoid edge effect. Two habitat types: hedgerow and winter wheat.
Time	Days	Daily energy expenditure simulated; simulations lasted for 20 years.
Exposure / effects	Repeated exposure Varying concentrations	Exposures were standardised using risk quotients (RQ; exposure divided by LD50or LC50). 4 scenarios (AllExposed-nonTK; AllExposed-TK; Spatial-nonTK; Spatial-TK) and compared to conventional risk assessment RQ without TK or IBM.
Abiotic environment	Food limitation	Not really food limitation, but depending on the habitat choice, energy input and pesticide uptake is modelled. Landscape designed as a torus (periodic boundary conditions).
Biotic environment	Intraspec. competition Predation	These processes impact individual growth and mortality.
Individuals	Energy budget Stochastic	Mice differ stochastically in their susceptibility to toxicity level, litter size and time of first reproduction. Other features such as maximum life span, gestation and lactation duration are constant. Individuals are characterized by the state variables sex and age that determines the life stage (pup, juvenile, adult).
Populations	Individual-based	

Criteria	Categories	Comments
Calibration	-	No calibration with field data, but structure similar to Wang and Grimm (2010) and Topping et al. (2003).
Programming language	Netlogo 4.1	Free software.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Pattern-oriented modelling with published field data on mice population without pesticide in Liu et al. (2013).
Sensitivity analysis	Yes	Robust to changes in mortality related parameters, but sensitive to changes in parameters related to reproduction.
Uncertainty analysis	Yes	The 4 scenarios (AllExposed-nonTK; AllExposed-TK; Spatial-nonTK; Spatial-TK) were compared together with differences in feeding-pattern; also conventional risk assessment with all 3 hypothetical pesticides differing in elimination half-lives. Authors found that for the exposed subpopulation including either spatial choice or TK reduced the RQ by 37–85%, and for the total population the reduction was 37–94%. However spatial choice and TK together had little further effect in reducing RQ. The reasons for this are that when the proportion of time spent in treated crop (PT) approaches 1, TK processes dominate and spatial choice has very little effect, and conversely if PT is small spatial choice dominates and TK makes little contribution to exposure reduction.
Documentation	Scientific publication	TRACE standard and ODD.

Assessment

Criteria	Description
Strengths	TK-model integrated into spatially-explicit IBM.
Theoretical uncertainties	Scenario comparison does to some extent not make sense: Comparing a scenario where all mice are exposed and a scenario with only 40% exposed mice (due to movement and feeding behaviour) logically leads to the conclusion that the impact on the population is lower when spatial behaviour is considered and that lower-tier RA is overprotective.
Empirical uncertainties	Could be high; based on few publications.
Parametric uncertainties	Could be high; feeding patterns (and thus exposure of population) are based on assumptions.
Temporal uncertainties	Biotic and abiotic environment was constant over 20 years and thus no additive effects of other processes (food limitation, higher predation levels) stemming from interannual variance were taken into account.

Criteria	Description
Conclusions	The wood mouse model is simpler than the ALMaSS approach (see section 1.3.4.8) but somewhat more complex and detailed than eVole (see section 1.3.4.6), e. g. due to the inclusion of nesting sites and the simulation of farming practices such as crop rotation. It has a high potential for application risk assessment, and some validation was done for predicted population dynamics without pesticide exposure. The authors concluded with their model results that lower-tier risk assessment is overprotective, which is a critical statement, given the lack of any validation on pesticide effects with independent data and theoretical uncertainties in their study.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Liu et al. (2013)	Wood mouse (*Apodemus sylvaticus*)	Generic	Original publication
Liu et al. (2014)	Wood mouse	Generic	Extension for TK and further improvement.
Bednarska et al. (2013)	Generic	Generic	Description of the incorporated TK module
Model applications			
Schmitt et al. (2015)	Wood mouse	Hypothetical fungicide	Comparison with 2 other mouse population models (ALMaSS, eVole). Similar predictions of risk.

1.3.4.8 **ALMaSS (Topping et al. 2003)**

ALMaSS (Animal, Landscape and Man Simulation System) is probably the most complex spatially explicit IBM that is currently available for the risk assessment of pesticides and comprises detailed mapping, weather, farm management, and vegetation growth. The landscape model has been first described with an application to voles in Topping et al. (2003). Each vegetated area has its own growth model, and in the case of farmed areas, management is modelled in detail. Animal models are agent-based, designed using the state/transition concept, and are rule-based. Each animal may interact with others and directly with its local environment. ALMaSS is modular and flexible so that it can be seen rather as a population modelling framework than a single model, and various models for different species have been created in which the details of the conceptual model may vary. Simulations of crop diversity and rotation demonstrate significant effects of spatial and temporal heterogeneity on population sizes, population fluctuations and landscape permeability. These two factors interact and thus different responses to temporal factors occur at different levels of spatial heterogeneity.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	
Questions / processes	Others	The original 2003 publication (EcoMod) had as purpose a predictive tool for answering policy questions regarding the effect of changing landscape structure or management on key animal species in the Danish landscape. Meanwhile, it is used for pesticide ERA.
Environmental domain	Terrestrial	
Taxon specificity	Generic, customizable	Generic; applications available for skylark (*Alauda arvensis*), roe deer (*Capreolus capreolus*), a ground beetle (*Bembidion lampros*), a linyphiid spider (*Erigone atra*), and the field vole (*Microtus agrestis*) (Topping et al. 2003, Jepsen et al. 2005, Topping et al. 2013, Topping et al. 2014); brown hare *(Lepus europaeus*) (Topping et al. 2010b, Topping et al. 2015).
Toxicant specificity	Generic	
Application	Applied in studies	76 citations in Web of Science; 131 in google scholar (Topping et al. 2003) (by June 2016)
Public availability	Source code	The project can be joined here: https://ccpforge.cse.rl.ac.uk/gf/project/almass/

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density	Landscape (land use type translated into habitat quality), farms (unit, type and crop husbandry planning; decide on crop rotation and hence land use type and habitat quality) and animals

Criteria	Categories	Comments
Endpoints	Population size Reproduction Recovery time Spatial distribution	Animal population; field vole: total population size; skylark: adults plus juveniles, the number of breeding pairs at the end of June each year, and the total number of fledglings produced that reached emigration age (i.e., total annual reproductive output); in later publications also spatial distribution of species (Jepsen et al. 2005) as well as time to recovery as population return to within 5% of control population densities (Topping et al. 2014).
Space	Grid cell	Vector map to grid cell conversion based on flyweight procedure. Each land use type has its own attributes and behaviours, e.g. roadside verges are subject to mowing whereas unmanaged grasslands are not. Vegetation growth is considered and crop rotation (management) is explicitly modelled. From this, habitat quality is derived. 1 m² can be modelled with an extent up to 100 km².
Time	Days	Time step = day, total duration can differ: simulation of 200 years (field vole application), 55 years incl. weather cycle (skylark application).
Exposure / effects	Toxicant mixtures	Field vole: No exposure, just landscape management. Four scenarios shared a common landscape structure, but differed in the farm management. Skylark: Scenarios considered the simple case of standard pesticide usage (P1) and zero pesticide usage (P2). In the case of monoculture crop scenarios (see below), a worst-case pesticide application was also evaluated whereby all farmers will apply a single standard dose of insecticide simultaneously. Varying plot exposure in spatial context (Topping et al. 2014) with exponential decay function.
Abiotic environment	Food limitation Temperature	Food limitation depending on habitat quality (field vole); temperature egg development and food limitation: scenarios simulate indirect effects of herbicide and insecticides on skylarks via a reduction in arthropod food availability (skylark), on beetles and spiders the insecticide is causing direct mortality.
Biotic environment	Intraspec. Competition Predation Parasitism	Species dependent, e.g. mortality through predation, starvation, infanticide (field vole); egg mortality, starvation, predation, migration mortality (skylark); included disease and density-dependent mortality (brown hare).
Individuals	Stochastic	Individual animals; the main processes growth, reproduction, mortality as well as dispersal and territoriality are based on habitat quality; density-dependence hence is an emerging property based on habitat quality and territoriality. Each of the main processes has many vole-specific sub-processes, e.g. reproduction consisted of maturation, mating, giving birth and lactation. The skylark model has different processes (e.g. egg development depending on weather conditions) and additionally cognitive attributes (memory of mates, offspring, geographical locations, nest location). Table 1 in Jepsen et al. (2005) gives good overview over processes per species.

Criteria	Categories	Comments
Populations	Individual-based	Population dynamics are an emergent property of the processes happening at the individual level.
Calibration	Field data	Used published field data for parameterization, whenever possible from Denmark.
Programming language	C++	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data Field data	Partly; Field vole: used visual interface and vole ecologists to verify plausibility of results as well as pattern-oriented modelling POM (Topping et al. 2012). Other species: no validation; brown hare and skylark (Topping et al. 2010b, Topping et al. 2013): 13 year's time series of population dynamics.
Sensitivity analysis	Yes	Habitat quality threshold HQT was sensitive parameter (field vole); extraction rate ER of food from habitat and minimum territory quality MTQ (skylark).
Uncertainty analysis	No	
Documentation	Website Scientific publication	Documentation well structured, follows ODD with own ODdox standard. Website with demo. Interface with R to import and create landscape layers (package 'ralmass' on github).

Assessment

Criteria	Description
Strengths	Full flex version IBM including the spatial component in a realistic way; focus on population dynamics; interactions between spatio-temporal environmental factors and the study organisms; inclusion of basic principles at the individual scale; integration of ERA at the landscape scale.
Theoretical uncertainties	Drift of pesticides to neighbouring patches not considered. Habitat suitability classes / energetic contents derived from land use types based on expert knowledge; many assumptions made to aggregate processes to a higher level.
Empirical uncertainties	There might be uncertainties in the life-history processes. Published field data and expert opinion were used.
Parametric uncertainties	Community level not modelled (e.g. interspecific interactions, trophic cascades); only single species with interspecific (mortality due to predation) or intraspecific (density regulation) interactions. Effect of pesticide modelled as increase in mortality and reproductive depression. Fecundity reduction is also an emergent property due to changed habitat suitability (= food availability) in models considering energetics.
Temporal uncertainties	No chronic effects on survival and reproduction after pulse exposure assumed.
Conclusions	The only model of this kind that deals so flexibly with landscapes, farm management and agent-based models. ALMaSS studies emphasize the need for greater focus on animal ecology in risk assessments.

Publications

Citations	Taxa	Chemicals	Comments
		Model description	
Topping et al. (2003)	Generic (here field vole example)	None	General presentation of the model with application to field vole.
Topping et al. (2010b)	Brown hare	None	Pattern-oriented testing and model development together with Oddox presentation with source-code documentation; hare model uses fundamental principles of energetics.
Topping et al. (2012)	Field vole	None	Shows complexity of fitting models to data.
Topping et al. (2013)	Skylark	None	Pattern-oriented testing and model development together with Oddox presentation with source-code documentation.
		Model applications	
Topping and Odderskaer (2004)	Skylark	Pesticides	Insecticide (Cyperb at a dosage of 0.25 L ha-1), herbicide (EK480 at a dosage of2Lha-1), fungicide (Tilt Turbo at a dosage of 1 L ha-1). Toxicant works via food reduction in combination with weather uncertainty and land management; for skylarks, metabolics (energy uptake and loss) have been explicitly modelled.
Jepsen et al. (2005)	Various	None	Same as Topping et al. (2003) paper with more species and spatial distribution, including carabid beetle (*Bembidion lampros*), a linyphiid spider (*Oedothorax fuscus*), a small farmland bird (skylark, *Alauda arvensis*), a small mammal (field vole, *Microtus agrestis*) and an ungulate (roe deer, *Capreolus capreolus*).
Topping et al. (2005)	Skylark	Pesticide	A comparison of a non-spatial IBM with ALMaSS handling and outcome; shows advantage of ALMaSS flexibility.
Dalkvist et al. (2009)	Field vole	Fungicide (vinclozolin)	Pesticide with complex long-term effects such as epigenetic transmission of reproductive depression. Vole ecology and behaviour were at least as important predictors of population-level effects as toxicology.
Topping et al. (2014)	Carabid beetle (*Bembidion lampros*), a linyphiid spider (*Oedothorax fuscus*)	Insecticide	Plot experiments for toxicant exposure. Importance to consider the large-scale impacts, not only local plots when assessing risk.

Citations	Taxa	Chemicals	Comments
Topping et al. (2016)	Brown hare	Insecticide	Higher Tier ERA of a fictitious endocrine disruptor; realistic landscapes compared; The model includes internal and external toxicokinetics (TK) in terms of the varying rates of ingestion of the pesticide, and the process of elimination within the hare. The internal TK are represented by a single compartment model assuming a percentage elimination rate per day. External TK is determined by the feeding behaviour of the hare and ultimately by the time spent feeding from contaminated areas, and the concentration of pesticide on vegetation. The study indicates that prediction of a reasonable worst-case scenario is difficult from structural, farming or population metrics; rather the emergent properties generated from interactions between landscape, management and ecology are needed.
Topping et al. (2015)	Carabid beetle (*Bembidion lampros*), a linyphiid spider (*Oedothorax fuscus*)	Insecticide	New area of landscape ecotoxicology; Pesticide stressors are simulated as changing spatial and temporal concentrations, based on spraying regimes and environmental fate of the active substances.

1.3.5 Empirical Models

1.3.5.1 Habitat Suitability Models (HSM, Chow et al. 2005)

Habitat suitability models (HSMs) are mainly used in conservation biology to predict species distributions through the modelling of proper environmental variables in space and time. HSMs are not mechanistic but statistical (or empirical) models that are trained with spatial environmental data on the occurrence of a species and on potentially relevant environmental variables. Predictions of HSMs therefore do not depend on profound prior knowledge of population processes but on environmental and species' distribution data (Thuiller and Münkemüller 2010).

Chow et al. (2005) developed a habitat suitability model to assess the environmental risk of contaminants for racoons using data from 13 radio-collared individuals. This probability resource selection model was implemented using knowledge of the spatial distributions of contaminants, an animal's home range, and spatial extent of the waste site. The exposure to a raccoon at a location is computed as a function of body weight, ingestion rate of media, and the concentration of contaminants within the media. The total exposure to racoons foraging at a waste site was modelled as a function of the ratio of waste site area to home range area weighted by the probability of the animal occurring within the area defined by its hypothetical home range. The contaminant exposure is a modified exposure estimation based on the work of Sample and Suter (1994), in which the proportion of the contaminated area that is suitable for the animal's use is replaced with the probability derived from the resource selection model.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific / Regulatory	Authors intended this paper as a contribution to the risk assessment for endocrine disruptors in fish.
Questions / processes	Body burden	Habitat suitability model based on resource-selection function combined with Gaussian plume to model risk exposure; aims at being a general framework for predicting contaminant exposure.
Environmental domain	Terrestrial	
Taxon specificity	Taxon-specific	Raccoon (*Procyon lotor*)
Toxicant specificity	Generic	Contaminated sediments; contaminants such as U, Ni, Al
Application	Little-known	
Public availability	-	

Variables and Parameters

Criteria	Categories	Comments
Entities	Body residues	The exposure to a raccoon at a location is computed as a function of body weight, ingestion rate of media, and the concentration of contaminants within the media. Average home range size needed for calculations.
Endpoints	Risk of exposure	

Criteria	Categories	Comments
Space	Grid cell	Hexagonal grid with 100 m side lengths extrapolated to approx. 800 km².
Time	Static model	
Exposure / effects	-	A lumped value of the contaminated media consumed was assumed to be 3.5 mg/kg/day for potential uptake of the species.
Abiotic environment	None	The resource selection function contains the following variables: habitat area, number of wetlands within the core area, distance to water, class landscape metrics
Biotic environment	None	No dynamical model.
Individuals	-	Averaged via hypothetical home ranges.
Populations	-	Indirectly assessed via hypothetical home ranges distributed across the landscape.
Calibration	Field data	Only the resource selection function has been fitted to field data.
Programming language	Visual Basic	Implemented as a dynamic linked library (DLL) in Environmental Systems Research Institute (ESRI ©) ArcMap (the GIS used by the DOE) using Visual Basic

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	No	
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Resource selection function (RSF) to detect places of foraging and high likelihood of contaminant uptake.
Theoretical uncertainties	No biological processes modelled; risk of exposure is a probability density function based on MC simulation; however, this is a statistical (static) model.
Empirical uncertainties	No biological processes modelled.
Parametric uncertainties	No biological processes modelled.
Temporal uncertainties	Static in time; not a dynamical model.
Conclusions	The idea of overlaying maps with most likely foraging places and contaminant presence is good; however, this is a static model and thus not useful for other species/ systems in ERA. To build proper RSFs from telemetry data is an extensive, data-hungry research field in itself, and therefore, the study is not easy to repeat for other species in different locations.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Chow et al. (2005)	Raccoon (*Procyon lotor*)	Contaminants such as heavy metals	

1.4 Community Level Models

1.4.1 Discrete Models

1.4.1.1 Model for Parasite-Host Interactions (Waage et al. 1985)

Waage et al. (1985) provide an example of a simple and general difference equation model for two non-staged populations that are connected. This model proceeds in discrete steps of generation times and has been used to simulate coupled parasitoid-host interactions, which is a typical application of this type of models (Soetaert and Herman 2009). Insect-induced mortality is considered in two different ways: Mortality acts between parasitism and reproduction or between reproduction and parasitism. Insecticide mortality acts in a density-independent manner. The model is limited to systems where parasitoids regulate the pest population.

General Properties

Criteria	Categories	Comments
Biological level	Population	
Model purpose	Scientific	To understanding the range of possible ecological interactions between pest, natural enemy and pesticide.
Questions / processes	Population recovery	Model focuses on population growth and mortality rates with function giving the proportion of hosts escaping from parasitism and parasitism rate. Importance of timing of spraying analysed relative to pest and parasitoid life histories.
Environmental domain	Terrestrial	
Taxon specificity	Generic	With application to spruce budworm in Canada and DDT spraying.
Toxicant specificity	Generic	
Application	Little-known	38 citations (Web of Science)
Public availability	-	

Variables and Parameters

Criteria	Categories	Comments
Entities	Population density	Population of host and parasite considered.
Endpoints	Population size Recovery time	Relative population levels.
Space	No spatial context	
Time	Generation times	Generic on generations.
Exposure / effects	Repeated exposure	
Abiotic environment	-	
Biotic environment	Parasitism	

Criteria	Categories	Comments
Individuals	None Homogeneous	Population level deterministic model.
Populations	Logistic growth	
Calibration	-	
Programming language	-	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	
Sensitivity analysis	No	Not a true sensitivity analysis, but different scenarios of mortality rates (e.g. different assumed levels of host susceptibility) and insecticide mortality timing. Important is effectiveness of the parasitoid in depressing the equilibrium pest population. Pest systems with less effective parasitoids will also show less resurgence when the timing of insecticide application causes parasitoid mortality.
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Simple mathematical model with general insights into patterns and dynamics of host-parasite interactions.
Theoretical uncertainties	Since it is a simple model based on simplified assumptions, the modelled system is well understood.
Empirical uncertainties	Many, as this is a simplified system.
Parametric uncertainties	Many, as this is a simplified system.
Temporal uncertainties	Many, as integrated over generation times.
Conclusions	Interesting paper to understand dynamics of mortality regimes on models with overlapping generations; not useful for application to real systems and data.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Waage et al. (1985)	Generic	Generic	

1.4.2 Continuous Models

1.4.2.1 TK Model for Aquatic Bioaccumulation (Arnot and Gobas 2004)

The food web model of Arnot and Gobas (2004) has been constructed to simulate the bioaccumulation of organic chemicals in different trophic levels of aquatic ecosystems. It is actually only a fate model because no effects of the body burden are calculated. However, the model shares many features with other community models and could be easily linked with an external effect module for effects at the species (represented by one individual) level. The model records the amount of toxicant on its way along the food chain, but does not keep mass balance of biomass (decomposition and nutrient cycling is not modelled). Each trophic guild (algae, phytoplankton, and macrophytes, zooplankton and small pelagic invertebrates, benthic invertebrates, water column filter feeders, small juvenile fish, medium sized fish, larger upper-trophic fish) is represented by one individual (compartment) of a typical species in the ecosystem that are connected via ordinary differential equations (ODEs). The model includes different possibilities of exposure to chemicals which are taken up through diet, directly from the water column, or by contact with pore water. The individuals are described as biomass which grows temperature-dependent through the ingestion of prey or through photosynthesis (algae, phytoplankton). The model has been parameterized with ecological field data from three North American lakes and validated with observed bioaccumulation factors (BAFs) for several organic chemicals in the lakes.

General Properties

Criteria	Categories	Comments
Biological level	Community / Food web	
Model purpose	Scientific / Regulatory	Risk assessment and environmental toxicological research of organic bioaccumulating chemicals.
Questions / processes	Others Body burden	Bioaccumulation of organic chemicals in a food-web. Site-specific estimates of chemical concentrations and the associated BAFs, BCFs and BSAFs.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	35 species from aquatic macrophytes, algae, phytoplankton, zooplankton, invertebrates and fish.
Toxicant specificity	Generic	Hydrophobic organic chemicals in aquatic ecosystems.
Application	Established in science	138 citations in Web of Science.
Public availability	Source code	Equations.

Variables and Parameters

Criteria	Categories	Comments
Entities	Growth	In case of modelling a steady-state, a constant growth rate is assumed, so growth is represented by a constant fraction of the body weight of an organism.
Endpoints		BAF (bioaccumulation factor) and BSAF biota-sediment accumulation factor).

Criteria	Categories	Comments
Space	No spatial context	
Time	Days	Timestep = 1 day.
Exposure / effects		The model can simulate a steady-state scenario in the case of chemicals with fast exchange kinetics, which are reaching steady-state relatively fast. Uptake of chemicals through diet, directly from the water or through exchange with pore water. Toxicant is eliminated from the organism via egg deposition or sperm ejection, metabolic transformation, growth dilution, and gill ventilation.
Abiotic environment	Temperature Water quality	Temperature, the degree of oxygen saturation in the water column, and organic carbon concentrations (POC & DOC) control algae growth and available food and temperature growth rates of the other species.
Biotic environment	Predation	Feeding on organisms in lower trophic levels.
Individuals		Individuals and population size in terms of individual numbers are not modelled.
Populations	Energy Budget	Biomass of one organism representing a trophic guild.
Calibration	Laboratory data Field data	Bioenergetic parameters are calibrated with laboratory and field data (freshwater species in three North American lakes).
Programming language	-	Programmed in Excel spreadsheets.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Comparison of the model outputs with observations of independent data for BAFs of organic chemicals in three North American lakes.
Sensitivity analysis	No	
Uncertainty analysis	Yes	Comparison of predicted model outputs with independent observation data and analysis of the overall error (model parameterization error, errors in model structure, analytical error in observed data, natural, spatial and temporal variability in observation data).
Documentation	Scientific publication	Well-structured scientific publication.

Assessment

Criteria	Description
Strengths	Model structure can compensate parameterization errors for the feeding rate and dietary uptake efficiency (often uncertain) and still provide correct BAF predictions. Simple model with requiring relatively little input for parameterization. Validation with independent data sets and application to independent chemicals.

Criteria	Description
Theoretical uncertainties	Every trophic guild is represented by one organism, so individual variability is not considered. Population effects and life cycle traits are not included (different age stages can be incorporated). No spatial context and environmental fate considered.
Empirical uncertainties	Amount of field observation data required for parameterization.
Parametric uncertainties	Exchange of non-ionic organic chemicals is described with one single equation for all aquatic species. The aqueous uptake clearance rate constant (the rate at which a chemical is absorbed from the water) is assumed to be identical for all species as function of the ventilation rate. Dietary uptake rate is difficult to calculate. Generalized growth rate equation is used when no observed growth rates are available. Assumption of one organism behaving representative for the whole population and one species representative for the whole guild.
Temporal uncertainties	Food web is assumed to be closed, without input from outside.
Conclusions	Modelling approach useful for risk assessment of bioaccumulating substances. Model outputs showed good fit with field observation data in most cases. Further development is needed to optimize the predictions.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Arnot and Gobas (2004)			Main publication reviewed here.
Gobas (1993)			First publication on the model approach.
Model applications			
Gobas and Arnot (2010)			Application to bioaccumulation of biphenyls in San Francisco Bay, California.

1.4.2.2 ODE Model for Freshwater Communities (De Laender et al. 2007)

De Laender et al. (2007) present a comparatively simple freshwater community model that uses a new approach of calibration. The model is not calibrated with time series data from a specific ecosystem, but with default values representing generic ecological concepts and seasonal events. The model is built up by three objects (phytoplankton, macrophytes, zooplankton), each describing increase or decrease in population biomass (growth) by differential equations. The processes considered by the model are photosynthesis, respiration, excretion, mortality, sinking and grazing by zooplankton for phytoplankton populations and grazing on phytoplankton and detritus, defecation, respiration, excretion and mortality for zooplankton populations. The objects can be used for describing a number of different populations. Laboratory test results (e.g. EC50, LC50) from individual tests are used for describing the toxicity of chemicals.

General Properties

Criteria	Categories	Comments
Biological level	Coommunity / Food web	
Model purpose	Scientific	Risk assessment of chemicals.
Questions / processes	Effect propagation	Predicting toxic effects of chemicals on populations from different planctonic categories (macrophytes, phytoplankton, zooplankton)
Environmental domain	Freshwater	
Taxon specificity	Generic	Generic use for fresh water species of phytoplankton, macrophytes and zooplankton.
Toxicant specificity	Generic	Application to copper, but use for other chemicals is possible if laboratory data is available.
Application	-	
Public availability	-	

Variables and Parameters

Criteria	Categories	Comments
Entities	Mortality Growth	Mortality increases while suboptimal temperature conditions or caused by exposure to toxic chemicals. Otherwise mortality is only described by the intrinsic mortality rate for every population. Growth is described by increase in population biomass.
Endpoints	Population size	Population biomass is observed for each simulation run and the average compared to the average biomass of a control run (RD). At the end NOECs are calculated.
Space	No spatial context	
Time	-	
Exposure / effects	Pulse exposure	A toxicant concentration is given as a default value and the effects on mortality rate or photosynthesis are calculated using concentration-response functions.

Criteria	Categories	Comments
Abiotic environment	Food limitation Temperature Light	Photosynthesis of phytoplankton population depends on temperature and light conditions. Suboptimal temperature enhances also zooplankton mortality.
Biotic environment	Predation	Feeding from zooplankton species on phytoplanktonic populations.
Individuals	-	No modelling of individual numbers or individual effects.
Populations	Biomass	Populations are described by their total biomass.
Calibration		Calibration by formulating differential equations for generic ecological concepts and dynamics. No calibration with field observation data for a specific site.
Programming language	Java	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Microcosm data	Comparison of predicted RDs and NOECs with values, calculated from observed microcosm data for six concentration-levels of copper sulphate with six planktonic freshwater species. Most of the time qualitatively correct prediction of biomass dynamics.
Sensitivity analysis	-	
Uncertainty analysis	Yes	A Monte-Carlo approach is used during the simulation run to represent variability in single-species toxicity test results.
Documentation	Scientific publication	Well-structured model description in a scientific publication.

Assessment

Criteria	Description
Strengths	Only singe-species toxicity test data (LC50, EC50), which are usually anyway measured for risk assessment are required as input parameters. Correct prediction of RDs and NOECs observed in a microcosm study. Abiotic factors influencing model species (temperature, light) are considered as well as diversity in individual sensitivity to toxicants.
Theoretical uncertainties	Density dependent effects are not considered. The food web is simple (e. g. fish missing), with just a few trophic levels, so that only predation on phytoplankton is assessed and no predation on zooplankton. Nutrients (P, S, POM and others) are not modelled in detail, which would increase relevance of model predictions for field conditions.
Empirical uncertainties	Model not validated with field observation data.
Parametric uncertainties	No spatial context and no chemical fate considered (default values for chemical concentrations).
Temporal uncertainties	

Criteria	Description
Conclusions	Interesting approach especially because just little input data is required. Adding more trophic levels to the food web and a chemical fate module would increase the relevance of model predictions, but also uncertainty.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
De Laender et al. (2007)	Freshwater plankton	Copper	

1.4.2.3 ODE Model for Resistance (Becker and Liess 2015)

The model of Becker and Liess (2015) has been developed to analyse the effects of intraspecific and interspecific interactions (competition, predation) on the spread of a pesticide resistance allele in exposed populations. The model combines the Lotka-Volterra differential equations for predation and for interspecifc competition to simulate the growth of three subpopulations of sensitive, heterozygous and resistant individuals, together with a predator and an interspecific competitor population. Within one generation time, the subpopulations are mixed based on Hardy-Weinberg equilibrium, assuming random mating. The three subpopulations differ in their growth parameters (intrinsic growth rate r, carrying capacity K, and relative competitive strength c), which are affected by pesticide exposure. Depending on exposure conditions and parameterization, the susceptible or the sensitive phenotype will dominate or even replace the other phenotype after several generations. The process is considerably fostered through intraspecific competition (when carrying capacity is approached). Predation and interspecific competition decrease the amount of intraspecific competition and delay phenotype replacement.

General Properties

Criteria	Categories	
Biological level	Community / Food web	
Model purpose	Scientific	
Questions / processes	Resistance development Additional stressors	Show effects of biotic interactions on the development of susceptible and resistant subpopulations during and after pesticide exposure.
Environmental domain	Generic	
Taxon specificity	Generic	Parameterized for mosquitoes and *Daphnia magna*.
Toxicant specificity	Generic	Parameterized for chlorpyrifos exposure and resistance in the mosquitoes (*Culex quinquefasciatus*).
Application	Little-known	
Public availability	Source code	

Variables and Parameters

Criteria	Categories	Comments
Entities	Population size	Biomass pools for macroinvertebrate taxa, periphyton, fine and suspended particulate organic matter.
Endpoints	Population size	Allele and phenotype frequency of sensitive, heterozygous and resistant individuals.
Space	No spatial context	
Time	Generation times	Phenotype and allele replacement in generation times.
Exposure / effects	Repeated exposure	Effects are simulated by changing the intrinsic growth rate, the carrying capacity and / or the relative competitive strength of the sensitive and heterozygous subpopulation.
Abiotic environment	Food limitation	

Criteria	Categories	Comments
Biotic environment	Intraspec. competition Interspec. competition Predation	
Individuals	None	
Populations	Logistic growth	
Calibration	Laboratory data	Calibration with data from standard toxicity tests, life table response experiments, and selection experiments including intra- and interspecific competition and harvesting (artifical predation) with and without chlorpyrifos exposure.
Programming language	R	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	Predictions from the model that species interactions typically hinder the development of resistance has been qualitatively confirmed in a field study (Becker and Liess 2017). No quantitative validation.
Sensitivity analysis	Yes	Variations of r (intrinsic growth rate) and K (carrying capacity) had largest effects.
Uncertainty analysis	No	
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	The population genetics model has been calibrated with data from an extensive selection experiment. It is based on simple and well-established principles (Lotka-Volterra, Hardy-Weinberg).
Theoretical uncertainties	Resistance in a population is assumed to be based only on a single resistance allele; interactions of more alleles or even genes (mechanisms) are not considered. Assumption of random mating (Hardy-Weinberg equilibrium) may be unrealistic but represents as best guess.
Empirical uncertainties	Relative competitive strengths were not measured but indirectly estimated through calibration in search of best model fit.
Parametric uncertainties	The model was analysed with parameter settings for functionally recessive and dominant heritability of resistance. Settings for overdominance and underdominance were not analysed.
Temporal uncertainties	The carrying capacities are fixed over time (no seasonality); no phenotype exchange through migration.
Conclusions	The model demonstrates how interacting species can hinder the onset of resistance development. It requires only few parameters but these are difficult to parameterize for other situations.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Becker and Liess (2015)	*Daphnia*, *Culex*	Chlorpyrifos	

1.4.3 Individual-Based Models (IBMs)

1.4.3.1 IBM for Effects of Competition and Pesticides (Kattwinkel and Liess 2014)

Kattwinkel and Liess (2014) present a model of two species that compete for the same resources to investigate effects of interspecific competition on population recovery after pulsed pesticide exposure. The model is not applied to specific species but is rather an approach for assessing general ecological issues. One of the modelled species is sensitive and the other insensitive to a generic toxicant. The species are assumed to have similar life-cycle traits: Aging, mortality, maturation and reproduction are modelled, but reproduction is the only density-dependent trait. Predicted population recovery is compared to recovery when only the sensitive species is present. In the model, interspecific competition can largely delay population recovery and even drive the sensitive species to extinction after repeated pulse exposure.

General Properties

Criteria	Categories	Comments
Biological level	Community / Food web	
Model purpose	Scientific	Demonstrating the importance of interspecific competition for population recovery after pesticide exposure and investigating how certain reproductive traits influence population recovery.
Questions / processes	Population recovery	Investigation of recovery time under interspecific competition with a non-sensitive species.
Environmental domain	Freshwater	
Taxon specificity	Generic	Generic, but life-cycle functions and model parameterization are guided by that of *Daphniidae* species.
Toxicant specificity	Generic	No specific toxicant included. Assumption that one of two species is sensitive to the toxicant and the other insensitive.
Application	Little-known	
Public availability	Source code	Model source code is available in the supplemental data.

Variables and Parameters

Criteria	Categories	Comments
Entities	Growth Mortality Reproduction Population density	Life cycle parameters (age of maturity, survival probability, intrinsic number of offspring, time between reproduction events) for each individual are gained from normal distributions of the species-specific values. Juveniles age with each time-step until they reach their age of maturity, from then they reproduce with gaps defined by their time between reproduction. Mortality is composed of the individual survival probability and acute toxic effects. Reproduction is affected by the intrinsic number of offspring and density dependent effects.

Criteria	Categories	Comments
Endpoints	Recovery time	Population recovery is estimated with the Wilcoxon rank sum test for unpaired samples. Recovery is achieved with less than 10% deviation from the control simulation.
Space	No spatial context	
Time	Days	Timestep = 1 day.
Exposure / effects	Single peak exposure	Assumption of one species being sensitive to the toxicant and the other being insensitive. Contamination occurs after an initialization phase (500 days). Acute effects occur at a single time step when exposure takes place and the toxic effect is modelled with values from 0 - 1 (0-100% mortality). No explicit toxicant modelled.
Abiotic environment	-	Food limitation is modelled indirectly through density dependent reproduction (effect of competition).
Biotic environment	Intraspecific competition Interspecific competition	Inter- and intraspecific resource competition for is considered as density dependent reproduction.
Individuals	Stochastic	Parameter values of individuals (age of maturity, survival probability, intrinsic number of offspring, time between reproduction events) were taken from normal distributions of the species-specific values.
Populations	Individual-based	Sum of individuals of one species make up the population.
Calibration		No calibration since no specific species were described in the model application.
Programming language	Java	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No	No validation with independent data.
Sensitivity analysis	Yes	For sensitivity analysis all parameters were varied between 25% - 200% of the default values (survival probability per time step between 75% and 102%), while the other parameters were kept constant. The mean population density depending on the parameters and the time to recovery (± 10 % of the control simulation run) after 50% reduction through contamination were analysed and competitive strength and carrying capacity were found to have to strongest influence on the mean population density of the sensitive species. Population recovery after 50% mortality was most affected by survival probability of both species. Population density and population recovery showed both to be insensitiev towards age of maturity, number of offspring and days between reproduction.
Uncertainty analysis	No	

Criteria	Categories	Comments
Documentation	Scientific publication	Documentation follows the ODD protocol for IBM and is well structured and helpful.

Assessment

Criteria	Description
Strengths	Consideration of life-cycle traits and interspecific competition. Positive is also the inclusion of variability between individuals.
Theoretical uncertainties	Since there were not concrete species modelled no validation of the model outputs with independent data was performed, so it is not known if the model is able to make correct predictions. No spatial context is simulated and it is not clear how risk of exposure to toxicants can be included and simulated. (How can laboratory data (LC50, LOEL, PNEC etc.) used for modelling?). No explicit consideration of abiotic factors like temperature, food limitation, habitat, pH etc. and biotic factors like predation.
Empirical uncertainties	Required data for quantitative assessing how strong one species affects another (competitive strength) are mostly not available and difficult to specify.
Parametric uncertainties	No explicit modelling of competition for food, habitat etc. but summarising all these parameters in the density dependent effects. Assumption of only acute toxic effects.
Temporal uncertainties	
Conclusions	The model makes clear, that interspecific competition can be an important factor of species recovery time after contamination with chemicals. For use in risk assessment the model has to be applicated to concrete examples of species and toxicants and the model output has to be compared with independent observed data sets. For this, further model developing is necessary as well as gathering the required data. At this state, the model is not ready for risk assessment.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kattwinkel and Liess (2014)	Generic freshwater invertebrates	Generic	

1.4.3.2 **Eco-SpaCE (Loos et al. 2010)**

Eco-SpaCE (Loos et al. 2010) is an individual-based, spatially explicit fate and effects model of a terrestrial food web with three levels for plants and invertebrates (not modelled as individuals but only as biomass), small vertebrates, and top predators in river flood plains. In the model, a toxicant causes acute mortality if the body burden exceeds an individual-specific threshold drawn from a distribution around the user-provided LC50. Therefore, so far, no sublethal effects but selection for resistant individuals is considered. Body burden increases with ingestion and decreases with excretion, and bioaccumulation is considered. Individuals gain or lose weight according to an energy-balance of ingestion and costs for maintenance, growth and reproduction. Behaviour and energy investigation in different processes (growth, mating, reproduction) change with different life stages (depending on age) and seasons (growth and breeding season). Foraging of small vertebrates is modelled as a random walk within a home range. Juveniles establish new home ranges at random suitable ecotypes in defined distances of their mother's home range. If the preferred diet is not sufficient, predators switch to alternate prey (ingested food fractions of suboptimum diet increase). Predation probability is a function of the predator's prey preference and prey abundance. Stochastic flooding events kill non-flying individuals. Most parameter values are drawn from distributions (random variables). The main publication (Loos et al. 2010) provides a quantitative comparison of the modelled effects of toxicants (cadmium), flooding and ecological stress (starvation and predation) on survival in scenarios with different intensities of each stressor.

General Properties

Criteria	Categories	Comments
Biological level	Community / Food web	
Model purpose	Scientific / Regulatory	Risk assessment of chemicals.
Questions / processes	Effect propagation Population recovery	Quantitatively compare effects of toxicant exposure and other stressors on a community in a heterogeneous environment.
Environmental domain	Terrestrial	Developed for river floodplains.
Taxon specificity	Generic	Developed for plants and soil-dwelling vertebrates as food source, small mammals, and top predators such as owls. May be applied to different species.
Toxicant specificity	Generic	Applied to cadmium.
Application	Known in science	
Public availability	Stand-alone program	

Variables and Parameters

Criteria	Categories	Comments
Entities	Individuals Biomass	Individuals belong to top predators or small vertebrates (several species possible, respectively). Plants and invertebrates as food for small vertebrates is modelled as biomass compartment.
Endpoints	Population size	

Criteria	Categories	Comments
	Body burden	
Space	Grid cell	Grid with 25 m² quadratic cells. Consists of multiple layers describing ecotope, elevation (protection from flooding), toxicant concentration in soil, and standing biomass of food. Individuals can only exist in suitable ecotopes.
Time	Days	
Exposure / effects	Varying concentrations Repeated exposure	Contaminant concentration in food is a function of local contaminant concentration in soil (user input).
Abiotic environment	Food limitation Temperature Flooding	Stochastic flooding events kill all non-flying individuals in affected cells.
Biotic environment	Intraspecific competition Interspecific competition Predation	Biomass of plants and invertebrates changes seasonally according to a sinusoidal function.
Individuals	Energy budget Stochastic	Stochastic values for most individual parameters, such as the lethal body residue concentration LBR.
Populations	Individual-based Biomass	Biomass compartment for plant and soil-dwelling invertebrates as food for small mammals in each cell.
Calibration	Laboratory data Field data	LC50 values for small mammals taken from rats and mice in the applications. Behaviour parameters estimated from field and laboratory studies.
Programming language	C++ using EcoSim 2.3 code libraries.	Early version in Schipper et al. (2008) uses also Visual Basic.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Comparison of modelled abundance without toxicant exposure with literature data in Loos et al. (2010). In most cases considerable over- or underprediction of maximum population densities. Cadmium body burden was predicted reasonably (Schipper et al. 2008). The predicted effects of toxicity, food limitation and predation were not tested, but predation was likely underestimated.
Sensitivity analysis	Yes	Seven scenarios with different stressor intensities. No analysis for other parameters.
Uncertainty analysis	Yes	75 % CI calculated from six model runs for each scenario.
Documentation	Scientific publication	Comprehensive description of mechanisms and results.

Assessment

Criteria	Description
Strengths	The IBM considers intraspecific variation in many traits, many relevant biotic and abiotic stressors, bioaccumulation and spatial aspects such as home ranges. The model can be parameterized with data studied on individuals.
Theoretical uncertainties	No parasitism, though parasites can have serious effects on small vertebrate populations.
Empirical uncertainties	LC50 values taken from rats and mice, may be not appropriate for the simulated species of small mammals in the model applications (voles, moles, mice).
Parametric uncertainties	In Loos et al. (2010), predation levels not realistic because only a single predator species modelled which cannot adapt the size of its home range in the model. Same for food limitation. No sublethal effects of toxicants, no mixtures.
Temporal uncertainties	No realistic food shortage and increased energy demand in winter.
Conclusions	While various relevant aspects are simulated in detail, some important aspects of toxic effects such as sublethal effects, mixtures and food availability are modelled not in sufficient detail, therefore high uncertainties in prediction. Tests of predicted toxicant effects missing.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Loos et al. (2010)	Moles, voles, and owls (top predator)	Cadmium	Main publication of the model and application to moles, voles and owls in a floodplain of the Rive Rhine in The Netherlands.
Loos et al. (2008)	Generic	Generic	Project report with model description.
Model applications			
Schipper et al. (2008)	6 small mammals and 4 top predators	Cadmium	Simulation of cadmium accumulation and comparison with field data on cadmium concentration to study the relevance of spatial distribution of cadmium pollution to body burdens.
van den Brink et al. (2011a)	Mice, voles	Cadmium	Collection of data on cadmium bioaccumulation for testing.

1.4.4 Empirical Models

1.4.4.1 PERPEST (Van den Brink et al. 2002)

PERPEST (Van den Brink et al. 2002) is not a mechanistic, but an empirical community effect model. After entering a given pesticide, concentration and ecosystem type, the model infers the predicted risk from comparable mesocosm studies. Each concentration from a study in the internal data base is considered as a case; the effect size for each species of each case is recorded on a 5-category scale. A search algorithm selects cases similar to the entered scenario using case-based reasoning. The data base also contains TU values for the most sensitive standard test organism; this enables to predict effects of a given pesticide based on results from related pesticides. The predicted risk is the mean of the deduction from each suitable case, weighted according to the similarity of the case with the entered scenario.

General Properties

Criteria	Categories	Comments
Biological level	Community / Food web	Use of data from artificial ecosystems (mesocosms and microcosms).
Model purpose	Regulatory	
Questions / processes	Effect propagation Population recovery	The model intends to replace mesocosm studies by comparing a user-defined scenario with results from similar previous mesocosm experiments.
Environmental domain	Freshwater	
Taxon specificity	Taxon-specific	Variety of test organisms in different microcosm and mesocosm studies.
Toxicant specificity	Toxicant-specific	Hydrophobic organic chemicals in aquatic ecosystems.
Application	Known in science	First publication 24 times cited in Web of Science.
Public availability	Stand-alone program	The model can be downloaded for free at http://www.perpest.alterra.nl/

Variables and Parameters

Criteria	Categories	Comments
Entities	Other	Cases (data sets from mesocosm studies).
Endpoints	Population size Biomass Water quality	The model predicts the probability of being affected to a certain degree. The groups of endpoints to be affected can be selected by the user, but is typically population size.
Space	No spatial context	
Time	Months	No dynamic simulation. The scales for the endpoints comprise 5 categories which discriminate between transient (< 8 weeks) and long-term effects observed in the studies.
Exposure / effects	Repeated exposure Chronic vs. pulse Varying doses	The user can select a scenario with pulse exposure or chronic / repeated exposure.

Criteria	Categories	Comments
Abiotic environment	Food limitation Other	No explicit modelling of abiotic conditions, but the mesocosm studies typically include some abiotic stressors such as food limitation.
Biotic environment	Intraspec. competition Interspec. competition Predation Parasitism	No explicit modelling of biotic conditions, but the mesocosm studies typically include all types of species interactions in a community.
Individuals	Other	No explicit modelling of individuals.
Populations	Other	Weighted average populations size or biomass of all suitable cases.
Calibration	Mesocosm data	See main description of the model above.
Programming language	-	Unknown.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Multiple calibrations Independent data	The predicted effects of metabenzthiazuron on community metabolism and the population size of phytoplankton, zooplankton and macrophytes were compared with mesocosm results not yet included in the model. NOECs from the mesocosm corresponded with predicted 50 % probability of finding a clear effect. Predicted and observed effects were in reasonable but not precise agreement.
Sensitivity analysis	Yes	The search algorithm for the selection of similar cases was optimized by trying different weights and values for the selection criteria. and analysed. The results were most sensitive to the range of the TU, the maximum distance (for scaling) and the distance power.
Uncertainty analysis	Yes	For each model run, a 95% CI for the prediction is calculated from bootstrapping: Randomly selected single cases are used for prediction, instead of the weighted mean over all suitable cases.
Documentation	Scientific publication	Scientific publication and technical documentation for download on website.

Assessment

Criteria	Description
Strengths	Based on a large data base of microcosm experiments. Few parameters that have to be estimated. Uncertainty analysis for each model run included.
Theoretical uncertainties	
Empirical uncertainties	The predictive power of the model strongly depends on the quality of the mesocosm studies. If sensitive taxa with long generation times and low recovery potential have been underrepresented in the mesocosms the model will underestimate effects.

Criteria	Description
Parametric uncertainties	Inferring effects from different substances and taxa may be misleading as the toxicological profile can differ even between related compounds and taxa.
Temporal uncertainties	Short-term mesocosm studies are typically not conducted during winter, therefore no predictions for overwintering stages.
Conclusions	The model is not a dynamic simulation but infers effects from case studies. Therefore, it requires minimal theoretical understanding but a large base of high-quality empirical data. Compared to dynamic simulations the predictions are rather imprecise but have a low probability of being completely wrong.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Van den Brink et al. (2002)	Freshwater organisms	Various pesticides	First publication of the model.
http://www.perpest.alterra.nl/			
Van Nes and Van den Brink (2003)	Freshwater organisms	Various pesticides	Manual and technical description.
Model applications			
Van den Brink et al. (2006)	Freshwater organisms	Atrazine	

1.4.4.2 $SPEAR_{pesticides}$ (Liess and von der Ohe 2005)

$SPEAR_{pesticides}$ (Species At Risk, Liess and von der Ohe 2005) has been developed as a bioindicator for exposure and effects of pesticides to freshwater macroinvertebrates in small streams. However, the approach can be also applied as an empirical community effect model, similar to PERPEST (see above). Macroinvertebrate species were classified as being at risk (SPEAR) or not at risk (SPEnotAR) based on four traits: Physiological sensitivity to toxicants (average acute LC50), recovery potential through reproduction (generation time), recovery potential through migration, risk of exposure (existence of aquatic life stages during pesticide application season). The SPEAR index expresses the ratio of observed SPEAR individuals vs. SPEnotAR individuals at a site.; low values indicate pesticide effects. A regression of observed SPEAR vs. the overall pesticide toxicity in water samples (expressed as summed up or maximum toxic unit, TUsum or TUmax) has been performed across various European small streams. TUsum or TUmax (sum or maximum of concentration divided by the LC50 for the reference species *Daphnia magna*) quantifies the toxicity of a pesticide mixture in a standardized way. Effects were driven by the most toxic substance (TUmax) rather than the summed up toxicity of all pesticides found (TUsum) . Therefore, it is possible with the observed regression to predict long-term effects on the macroinvertebrate community composition from exposure to a pesticide with a given TU, if this pesticide is driving the overall toxicity (TU ≈ TUmax). Other versions of SPEAR are available for organic pollutants and salinity (applied in Australia).

General Properties

Criteria	Categories	Comments
Biological level	Community / Food web	
Model purpose	Scientific	
Questions / processes	Effect propagation	
Environmental domain	Freshwater	Established and validated for small streams.
Taxon specificity	Taxon-specific	Freshwater macroinvertebrates.
Toxicant specificity	Generic	Pesticides
Application	Established in science	
Public availability	Stand-alone program	SPEAR calculator is part of the Indicate software that is available for free download at https://www.ufz.de/index.php?de=38122

Variables and Parameters

Criteria	Categories	Comments
Entities	Life history traits Population size	Species are classified as SPEAR / SPEnotAR based on sensitivity and vulnerability (life history traits).
Endpoints	Community composition	SPEAR index (based on the ratio of SPEAR vs. SPEnotAR) quantifies changes in the community composition due to pesticide exposure. The indicator is largely independent from the effects of additional stressors.
Space	No spatial context	Presence of upstream recovery sites can be considered (separate regression for sites with upstream recovery area available).

Criteria	Categories	Comments
Time	-	No dynamic simulation. SPEAR predicts long-term effects from typical exposure patterns during the pesticide application season.
Exposure / effects	Repeated exposure Varying concentrations Toxicant mixtures	The model is calibrated to typical exposure scenarios in agricultural streams, i. e. repeated pulse exposure after run-off during the spraying season.
Abiotic environment	-	The model uses a regression obtained from field data and thus implicitly incorporates all stressors typically found in the field.
Biotic environment	-	The model uses a regression obtained from field data and thus implicitly incorporates all stressors typically found in the field.
Individuals	-	No explicit modelling of individuals.
Populations	-	The model does not predict effects for specific populations, but quantitatively predicts changes in the community composition of generic species groups SPEAR an SPEnotAR.
Calibration	Field data Laboratory data	Classification of species as SPEAR / SPEnotAR based on life history traits reported in various scientific literature. Establishment of SPEAR vs. TU regression with field data.
Programming language	-	Unknown.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	The model has been calibrated with data from numerous independent communities in the field. It has been applied in four different continents (Europe, Australia, Siberia, South America) and predicted similar effects on the community composition in all case studies. Hence validated in several field studies.
Sensitivity analysis	No	
Uncertainty analysis	Yes	R^2 of observed TU - SPEAR regression (typically around 0.5).
Documentation	Scientific publication	

Assessment

Criteria	Description
Strengths	Simple model for the prediction of the effects of toxicant mixtures under realistic scenarios in the field. The model uses extensive field data from numerous independent sources which represents a way of validation that has not been applied to any other reviewed model. The endpoint SPEAR quantifies the state of a community according to the protection goal of no permanent decrease in any population; any pesticide-induced population decrease results in a decreased SPEAR value.

Criteria	Description
Theoretical uncertainties	As generic species and toxic units are the basis of the model, effects cannot be attributed to specific species or to specific mode of actions.
Empirical uncertainties	As trait associations to various species are uncertain also the overall prediction includes species related uncertainty.
Parametric uncertainties	The model is not able to predict a risk for any specific population / species. It cannot be adjusted to properties of specific substances (except for LC50).
Temporal uncertainties	The model is not dynamic and was calibrated to typical scenarios observed in the field. It may not be applicable to untypical exposure, climatic or community scenarios.
Conclusions	The simple and empirical model implicitly includes all conditions relevant in typical field scenarios. Therefore, it is expected to give not highly precise, but comparatively reliable predictions for scenarios within its scope, with a low risk of underestimating real effects. Extrapolation to untypical scenarios (not covered by calibration data) is difficult. Extension to predict effects on ecosystem functions (leaf litter breakdown) is turning the model to an ecosystem model.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Liess and von der Ohe (2005)	Freshwater macroinvertebrates	Generic	Original publication.
Beketov et al. (2009)			Comparison of SPEAR/SPEnotAR classification at species and family level.
https://www.ufz.de/index.php?de=38122			Website
Model applications			
Hunt et al. (2017)	Freshwater macroinvertebrates	Generic	Application to streams in the Argentinian pampas.
Münze et al. (2015)	Freshwater macroinvertebrates	Generic	Link of the SPEAR indicator to ecosystem functions (leaf litter breakdown).
Orlinskiy et al. (2015)	Freshwater macroinvertebrates	Generic	Effects of upstream recovery sites on the SPEAR - TU relation.
Schaefer et al. (2012)	Freshwater macroinvertebrates	Generic	Link of the SPEAR indicator to ecosystem functions (leaf litter breakdown).
Schaefer et al. (2011)	Freshwater macroinvertebrates	Generic	Application to Australian streams.
Schletterer et al. (2010)	Freshwater macroinvertebrates	Generic	Application to Siberian streams.
Liess et al. (2008)	Freshwater macroinvertebrates	Generic	Application to different European countries.

1.5 Ecosystem Level Models

1.5.1 Freshwater Models

1.5.1.1 **AQUATOX (Park et al. 2008)**

AQUATOX (Park et al. 2008) is probably the most comprehensive aquatic ecosystem model available. The model is based on several hundred coupled differential equations that truly integrates the fate and effects of toxicants in various types of ecosystems that can be simulated, ranging from ponds to lakes, streams, rivers and estuaries. Development dates back to the first predecessor CLEAN in 1974 and is currently done by the USEPA. AQUATOX does not model individuals but changes in the overall biomass of biotic compartments (populations) and abiotic compartments (detritus), The model provides a detailed mechanistic description of abiotic and biotic processes such as stratification in lakes, seasonal changes in temperature and light, sediment transport in streams, nutrient cycling, and bioaccumulation in the food web. The food web can be flexibly set up with one or several species per guild or functional group (except for microorganisms that are not explicitly modelled). Deceived or excreted biomass passes several forms of detritus before it is re-mineralised to dissolved nutrients (C, N, and P) in the water column and assimilated again through photosynthesis. Toxicants are partitioned among the biotic and abiotic compartments through sorption and desorption, uptake (feeding and uptake at gills with respiration), depuration of organisms (excretion and defecation), and decay of biomass (detritus). Stochiometric functions keep mass balance of the modelled nutrients and toxicants in the system. Toxicants are also subject to photolysis, hydrolysis, volatilization, microbial degradation and biotransformation (with the possibility of forming metabolites with their own toxicity). Lethal and sublethal direct effects of up to 20 organic toxicants on the biomass of a population are simulated based on their internal concentrations within the biotic compartment. This is done with a unique TKTD module which requires only LC50 (and EC50 for sublethal effects) as ecotoxicological input for a given combination of toxicant and species. However, the module makes a number of generalising assumptions that can increase uncertainty in the model predictions. Direct effects of multiple toxicants are considered to be additive and can propagate in various ways through the food web. AQUATOX is not applicable to metals. Due to its high complexity, the parameterization of AQUATOX is challenging. The model provides various built-in scenarios that can be used as a starting point for the creation of own settings. Built-in libraries for properties of various sites, species and chemicals then facilitate the modification of a built-in scenario. To meet the notorious lack of ecotoxicological data for parameterization, AQUATOX can be linked to the WebICE application that estimates LC50 values for a given taxon and compound from regressions with related taxa and compounds from an extensive data base.

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	
Model purpose	Scientific / Regulatory	Support for risk assessment of organic toxicants and water quality management.
Questions / processes	Effect propagation Population recovery	Predicting the environmental fate (bioaccumulation) and effects of h of organic toxicants in various aquatic ecosystems. The model has been developed for the prediction of ecological response to proposed strategies in water quality management and environmental risk assessment.
Environmental domain	Freshwater	

Criteria	Categories	Comments
Taxon specificity	Generic	All guilds and functional groups of a freshwater community can be represented by several user-defined surrogate species.
Toxicant specificity	Generic	Up to 20 organic chemicals simultaneously (no metals).
Application	Established in science Applied for retrospective risk assessment	
Public availability	Stand-alone program	The model can be download for free at https://www.epa.gov/ceam/aquatox-32-download-page.

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	Each surrogate species represented by 1 biotic compartment for biomass and stored toxicants (but fish can be modelled using multiple compartments for age or size classes). Abiotic compartments include 8 types of detritus, the water column, and optionally inorganic sediment. Water column in lakes may be separated in epilimnion and hypolimnion. Several connected stream stretches can be modelled, each with the full set of biotic and abiotic compartments listed above.
Endpoints	Biomass Water quality Recovery time Body burden	As AQUATOX is very flexible and detailed, each state variable can be used as endpoint, depending on the question addressed. Algal and moss biomass can be converted to chlorophyll a content, and various biotic indices can be calculated for better comparison with monitoring data.
Space	No spatial context	User-provided volume, depth and surface area of the modelled water body affect processes such as sedimentation and light attenuation in the simulation. Several connected stream segments can be modelled which provides an implicit representation of space. However, distance and size of the stretches is not explicitly modelled, and the no spatial differentiation is made within the compartments of a site.
Time	Days Hours	Time step is 1 day be default. Can be changed to 1 hour for the simulation of diurnal O_2 fluctuations or rapidly degrading toxicants.
Exposure / effects	Chronic vs. pulse Varying concentrations Toxicant mixtures	Toxicants loadings from a user-specified driving variable enter the water column and are partitioned among the abiotic and biotic compartments via sorption and desorption, uptake through gills and feeding, excretion, defecation and decay of biomass.

Criteria	Categories	Comments
		Direct lethal effects are calculated based on the internal concentration, using a simplified TKTD module: First, a user-provided external LC50 is converted to an internal LC50 which decreases with increasing exposure time, based on physicochemical properties of the toxicant. Each time step, the internal LC50 for the current exposure time (duration of previous exposure to any concentration) is incorporated into a generic Weibull dose-response model. The toxicant-induced mortality from this model is compared to the highest mortality obtained during previous exposure. An excess in mortality is simulated as the actual mortality experienced during the given time step. This concept follows the logic that sensitivity of biomass increases (as LC50 decreases) with exposure time: X % of biomass that survived exposure in previous time steps will survive similar or lower concentrations also in the current time step, unless the LCx drops below these concentrations due extended exposure. The strength of direct sublethal effects is calculated from the same Weibull model, after the internal LC50 for the current exposure time has been multiplied with a sublethal:lethal ratio (default = 0.1). Direct sublethal effects include reduced photosynthesis, accelerated sinking of phytoplankton, reduced growth (in animals split into reduced consumption and reduced assimilation of consumed food), reduced reproduction, increased sloughing of periphyton and increased drift of invertebrates. Direct effects of multiple toxicants are considered to be additive.
Abiotic environment	Food limitation Temperature Light pH Dissolved oxygen	Biotic processes are affected by suboptimal levels of nutrients, temperature, light, pH and dissolved oxygen. Additionally, low levels of dissolved oxygen and high levels of ammonia cause lethal and sublethal toxic effects that are modelled similar to those of toxicants.
Biotic environment	Intraspecific competition Interspecific competition Predation	Growth of biotic compartments is limited by the availability of nutrients or prey (intraspecific competition), predation, and by competition with other biotic compartments for light, nutrients and prey (interspecific competition).
Individuals	None	Biomass compartments are not structured in individuals.
Populations	Biomass	Populations are described by their total biomass. Biomass of a fish population can be structured in size or age classes. Instantaneous loss of biomass due to spawning or the emergence of adult insects (that leave the simulated water body) can be modelled at given dates or water temperatures.
Calibration	Field data Mesocosm data	The built-in parameters for basic process rates have been mostly obtained from the literature and were rarely subject to calibration. The various applications of AQUATOX (some of which produced the built-in scenarios) used varying degrees of site-specific calibration with field or mesocosm data.

Criteria	Categories	Comments
Programming language	Pascal	Release 3 has been written using Pascal with the Borland Delphi 2007 development platform.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Validation of the predicted population development (mainly for algae) and of predicted bioaccumulation in some streams, lakes and estuaries. However, no validation of predicted direct or indirect toxicant effects on population development.
Sensitivity analysis	Yes	Modelled populations were particularly sensitive to parameters that describe the temperature-dependency of processes and to the water temperature itself, but also to biotic processes (consumption and respiration). Toxicant fate and effects were highly sensitive to log K_{OW}. Simpler food webs were more sensitive to toxicant-induced food web effects than more complex food webs.
Uncertainty analysis	Yes	A built-in automated uncertainty analysis can be run with user-defined parameter values and loadings being randomly drawn from Latin Hypercube sampling.
Documentation	Scientific publication Built-in help	Scientific publications, and a comprehensive documentation at the USEPA website. Built-in help functions to assist mode application.

Assessment

Criteria	Description
Strengths	Probably the most comprehensive ecosystem model available for the integrative simulation of fate and effects of toxicants within aquatic ecosystems. Includes numerous physical and ecological processes potentially relevant for risk assessment that are typically not explicitly considered in other effect models, such as biomagnification, and the potential of secondary exposure and changes in water quality and nutrient fluxes when contaminated organisms deceive and decay.
Theoretical uncertainties	As in most other effect models, the simulation of direct effects does not consider changes in susceptibility due to additional stressors or due to different life stages (except for fish). The modelled direct sublethal effects do not include many potential behavioural changes that may lead e. g. to exposure avoidance, increased susceptibility to predation or a change in feeding preferences. However, sublethal effects are potentially better represented than in most other effect models.
Empirical uncertainties	Due to several hundred parameters involved, parameterization of AQUATOX is very complex. Some parameters notoriously high uncertainty because they are subject to high variability in the field and typically not accessible to exact measurement (e. g. light extinction coefficients for periphyton, ecotoxicological data for non-standard species).

Criteria	Description
Parametric uncertainties	As populations are modelled as unstructured biomass compartments, no demographic effects can be modelled (except to some extent for fish which can be separated in age classes). Compartments are assumed to be well-mixed, ignoring potential spatial heterogeneity in pesticide exposure and other environmental conditions.
Temporal uncertainties	As no energy budget is modelled, delayed direct effects, such as an increased mortality of insects during pupation after exposure to sublethal concentrations in early larval instars, cannot be simulated.
Conclusions	AQUATOX has been successfully applied to describe the development of algae, fish and some invertebrates within their ecosystem context, as well as the fate and biomagnification of toxicants within aquatic ecosystems. The model has a high potential of mechanistically and realistically describing the propagation of pesticide effects and has been applied for retrospective risk assessment. E. g., the model was calibrated to specific rivers or lakes to understand the propagation of observed effects and to predict improvements in water quality due to mitigation measures. However, the built-in TKTD module for direct effects of toxicants requires various generic assumptions and is thus associated with high uncertainty. Predictions of AQUATOX on the direct and on the overall effects of toxicants have not been validated yet with independent observations, which is probably the most important reason why the model has not yet been applied for the prospective risk assessment of pesticides.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Park et al. (2008)	Freshwater organisms	Generic	Scientific publication on Release 3.
Park and Clough (2018)	Freshwater organisms	Generic	Technical documentation for Release 3.2.
Raimondo et al. (2010)	Freshwater organisms	Generic	User manual for WebICE 3.1.
https://www.epa.gov/ceam/aquatox			USEPA website for AQUATOX.
Model applications			
USEPA (2000)	Freshwater organisms	Various	USEPA model validation report.
USEPA (2001)	Freshwater organisms	Various	USEPA model validation report (addendum).
USEPA (2013)	Freshwater organisms	Generic	USEPA sensitivity analysis report.
Sourisseau et al. (2008)	Freshwater organisms	None	Application and validation with non-exposed artificial streams.
Zhang et al. (2013)	Freshwater organisms	PCBs	Calibration and retrospective risk assessment in Baiyangdian Lake (China).
Lombardo et al. (2015)	Freshwater organisms	alkylbenzene sulfonate, triclosane	Calibration to River Thames (as control scenario) and prediction of effects of added toxicants (no validation).

1.5.1.2 **CASM (Bartell et al. 1999)**

The Comprehensive Aquatic Systems Model (CASM) 2.0 from Bartell et al. (1999) is a complex ecosystem model that is based, on a set of differential equations for the increase and decrease in biomass of various freshwater species. Up to 60 representative producer and consumer populations of different guilds can be modelled. The model shares many features with AQUATOX, but is simpler, particularly because it incorporates no full mass balancing fate module for toxicants. Biomass growth of a modelled population is affected by nutrient and light availability (producers), prey availability (consumers), temperature, water quality, and population-specific growth parameters such as transition efficiency and natural mortality. Toxicants reduce the growth rates during exposure; the reduction due to a given concentration is estimated from a generic probit dose-response model with a user-provided LC50 for each species. CASM represents a family of similar ecosystem models based on bioenergetics (SWACOM, CATS, LERAM). All have been developed with USEPA and scientifically applied for multiple times, but (in contrast to AQUATOX) have no formal recommendation from USEPA.

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	Population growth and nutrient cycling in food webs.
Model purpose	Scientific / Regulatory	Risk assessment of chemicals.
Questions / processes	Effect propagation Population recovery	Direct and indirect effects of toxicants on the growth of populations in a food web, and on the resulting water quality in aquatic ecosystems.
Environmental domain	Freshwater	
Taxon specificity	Generic	Default data sets for representative producer and consumer species in Canadian streams and lakes; can be changed to arbitrary taxa. Decomposers are not explicitly modelled.
Toxicant specificity	Generic	
Application	Applied for retrospective risk assessment	Applied in ecosystem restauration studies and for non-regulatory assessment of toxicant effects.
Public availability	Stand-alone program	Can be requested from authors on website http://www.dsllc.com/modeling-tools/casm/

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	Population growth based on bioenergetics. Nutrient cycling in food webs is modelled with consequences on water quality.
Endpoints	Biomass Water quality	Biomass of each population, and state variables that define the water quality (e.g. DOM, DO) are reported.
Space	No spatial context	
Time	Days	Runs by default for 1 year.

Criteria	Categories	Comments
Exposure / effects	Chronic vs. pulse Varying concentrations Repeated exposure	The toxicant concentrations can be varied every day. Physiological interactions of toxicant mixtures not considered (additive effects only). LC50 based.
Abiotic environment	Food limitation Temperature Light Water quality	Cycling of C, N, P and Si. Nutrients and light limit the temperature-dependent growth of producers based on bioenergetic parameters. Temperature and light changes seasonally. Temperature and food limit the growth of consumer species.
Biotic environment	Intraspecific competition Interspecific competition Predation	Predation and food limitation due to intra- and interspecific competition.
Individuals	None	Biomass compartments are not structured in individuals.
Populations	Biomass Energy budget	Biomass of each producer and consumer population.
Calibration	Field data Laboratory data	Bioenergetics of temperate freshwater species from laboratory tests. Abiotic conditions and biotic interactions from field studies in Canadian freshwater.
Programming language	FORTRAN	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data Laboratory data	Predicted mean daily biomass of different guilds in simulations without toxicants were compared to observations in Canadian rivers. Simulations overestimated zooplankton and omnivorous fish, and underestimated benthic invertebrates, but fitted well to the other observed biomasses. In case studies, the predicted risk of field-derived concentrations of PCP, copper, diquat dibromide and mercury for total guilds was comparable to expectations from a probabilistic framework.
Sensitivity analysis	Yes	Model parameters were randomly varied for +/- 1% CV, assuming normal distribution.
Uncertainty analysis	Yes	Uncertainty of results due to uncertainty of entered LC50 is analysed with every model run. Monte-Carlo Test with alternate LC50 values based on normal distribution for the uncertainty of the predicted biomass reduction. Probabilities that the decrease of biomass exceeds different thresholds are reported.
Documentation	Website Scientific publication	Well structured, exhaustive description, following no specific guideline.

Assessment

Criteria	Description
Strengths	Prediction of risk incl. uncertainty analysis for each modelled population and for water quality parameters due to direct and indirect effects. Once a representative community has been parameterized, the model requires only acute LC50s for each new case. Site-specific information can be arbitrarily added if available.
Theoretical uncertainties	Life history of species not differentiated in different life stages, though sensitivity and vulnerability may vary with life stage. No chronic (delayed) effects of acute exposure. No variation in sensitivity of individuals.
Empirical uncertainties	Requires very much field data. Extrapolation of many biological parameters such as LC50, prey preferences or decomposition rates to related taxa and different water bodies is uncertain.
Parametric uncertainties	Assumes immediate response of populations, though long-living populations grow slower than those with short generations. The missing modelling of spatial context and environmental fate of toxicants ignores the possibility of heterogeneous exposure (or avoidance) in a heterogeneous environment. No explicit modelling of different decomposer taxa which may be also affected from toxicants and then affect water quality, food supply and mineralization. Simplification of populations to energy and nutrient budget may be only appropriate for large populations (and therefore low toxicant effects) where Allee effects and demographic stochasticity are negligible.
Temporal uncertainties	Considering the ecosystem as being closed ignores recovery through immigration or invasion by tolerant species.
Conclusions	Though initial parameterization for representative European freshwater communities is labour-intensive, the model offers the potential assessment of indirect effects of toxicants on many interacting species and ecosystem processes. It is thus potentially useful for the assessment of effect propagation from populations to communities and ecosystems. However, simplifying populations to energy and nutrient budgets ignores several relevant mechanisms that may affect the sensitivity of populations, such as chronic effects and individual variation in sensitivity. The potential effect size of such mechanisms and the specific risk for small populations due to demographic stochasticity should be assessed before use.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Bartell et al. (1999)	Freshwater species	Generic	An ecosystem model for assessing ecological risks in Québec rivers, lakes and reservoirs (original publication of CASM 2.0)
Bartell et al. (1986)			Comparison of numerical sensitivity and uncertainty analyses of bioenergetic models of fish growth.
Bartell et al. (1992)			
http://www.dsllc.com/modeling-tools/casm/			Website
Model applications			
Naito et al. (2003)			Application of CASM for aquatic ecological risk assessment of chemicals in a Japanese lake.
Amemiya et al. (2007)			Stability and dynamical behaviour in a lake-model and implications for regime shifts in real lakes.
Wu et al. (2010)			A risk-based decision model and risk assessment of invasive mussels.

1.5.1.3 **CATS (Traas and Aldenberg 1996)**

CATS (Contaminants in Aquatic and Terrestrial ecoSystems) from Traas and Aldenberg (1996) is an integrative fate and effects model family developed in collaboration with the Dutch environmental agency. Biomass pools for different functional groups are simulated based on coupled bioenergetic differential equations, and connected through a food web. Additionally, few abiotic compartments are considered for biomass and toxicants. Uptake of toxicants (including organic compounds and metals) into biotic compartments and bioaccumulation is modelled explicitly. Effects are calculated based on body burden, therefore bioaccumulation is considered. The model comprises 143 parameters; the complexity is thus simpler as compared to AQUATOX, but may be compared to those of CASM. Here we focus on applications to freshwater systems (see application examples below).

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	
Model purpose	Scientific / Regulatory	Risk assessment of chemicals.
Questions / processes	Effect propagation Population recovery	Study direct and indirect effects of toxicants in ecosystems under consideration of bioavailability and bioaccumulation.
Environmental domain	Generic	Mostly used for freshwater systems.
Taxon specificity	Generic	All types of communities.
Toxicant specificity	Generic	Organic compounds and metals.
Application	Applied for retrospective risk assessment	Applied in ecosystem restauration studies and for non-regulatory assessment of toxicant effects.
Public availability	Stand-alone program	

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	Biomass and toxicant pools for functional species groups and abiotic compartments.
Endpoints	Biomass Body burden Water quality	
Space	No spatial context	
Time	Days	
Exposure / effects	Chronic vs. pulse Varying concentrations Repeated exposure	No applications of toxicant mixtures, though modelling of mixtures with additive effects might be possible. Only acute effects.
Abiotic environment	Food limitation Temperature Light	

Criteria	Categories	Comments
	Water quality	
Biotic environment	Intraspecific competition Interspecific competition Predation	Food web model.
Individuals	None	
Populations	Biomass Energy budget	Bioenergetic growth curves for biomass pools.
Calibration	Field data	For numerous free parameters, random combinations of parameter values were compared to identify combinations that provide the best fit of predicted and observed endpoints.
Programming language	FORTRAN, ACSL	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data Mesocosm data	Insufficient comparison of predicted and observed body burden of aquatic invertebrates in application examples for CATS-2. No good accordance of predicted and observed values.
Sensitivity analysis	Yes	In some applications.
Uncertainty analysis	Yes	Monte Carlo simulation for prediction of toxicant accumulation is not automatically performed but was accomplished in application examples.
Documentation	Scientific publication	Comprehensive explanation of the model structure and example applications.

Assessment

Criteria	Description
Strengths	Integrative fate and effects model; high ecological realism, but needs only 1/3 of parameters compared to AQUATOX. Consideration of indirect effects, bioaccumulation, some abiotic conditions and effect propagation through changes in ecosystem functions.
Theoretical uncertainties	Direct effects modelled only as acute mortality, sublethal and chronic effects ignored.
Empirical uncertainties	Bioenergetics may highly vary in different environments.
Parametric uncertainties	No IBM or different life stages, therefore no variation within species considered. Aerial / terrestrial stages of amphibious invertebrates not considered in the reviewed applications to a freshwater system.
Temporal uncertainties	Mass-balance models assume mature ecosystems (no ongoing natural succession).

Criteria	Description
Conclusions	The model represents a promising compromise in the trade-off between excessive detail and ecological realism of an ecosystem model. However, effect predictions have been rarely tested with real data, and the existing validations generally showed low accordance of predicted and observed data.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Traas and Aldenberg (1996)	Grassland community, Freshwater community	Metals, TBT	Description of CATS-2 and application to metals in grasslands and to TBT in lakes. Very few data to test predicted TBT body burden, do not fit to predictions; predicted effects not tested. Risks of metals considered as probabilities that predicted body burden exceeds pre-defined thresholds. Validation missing.
Traas and Aldenberg (1992)	Grassland community	Metals	Description of CATS-1 and application to metals in meadows. No validation.
Model applications			
Traas et al. (1996)	Freshwater community	TBT	Application of CATS-2 to TBT in lakes. Few data for validation and no good accordance of predicted and observed TBT burden.
Traas et al. (1998)	Freshwater microcosms	Chlorpyrifos	Recovery of invertebrates in freshwater microcosms. Biomass dynamics and pesticide effects accorded roughly to observed time series, with high uncertainty.
Traas et al. (2004)	Freshwater microcosms	Chlorpyrifos	Modification called C-COSM to analyse effects of insecticides and nutrients and recovery in microcosms. Validation showed good prediction of fate and recovery, underestimation of fate of nutrients.

1.5.1.4 **Streambugs (Schuwirth and Reichert 2013)**

Streambugs (Schuwirth and Reichert 2013) provides a combination of classical food web modelling, metabolic theory of ecology (MTE) and ecological stoichiometry for risk assessment in streams. Mass balance with growth, respiration, and mortality is modelled; nutrients and oxygen are included for mass balance, but not modelled as state variables. To reduce the number of free parameters, bioenergetics of each species were fitted using allometric scaling according to the assumptions of MTE. Factors for growth, respiration and mortality, and several constants for the calculation of the basal metabolisation rate are assumed constant for all invertebrates. Predators feed on all smaller taxa, unless specific feeding behaviour is specified. Input of leaf litter as food source is included. Light-intensity for photosynthesis depends on season, shading and depth. The local carrying capacity of a taxon is modified based on its preferences for current, temperature and substrate. Pesticides and organic pollution affect sensitive species (classification according to $SPEAR_{pesticides}$ and saprobic index) depending on the concentration in the water.

General Properties

Criteria	Categories	
Biological level	Community / Food web	
Model purpose	Scientific / Regulatory	
Questions / processes	Others	Predict the community composition of macroinvertebrates in streams.
Environmental domain	Freshwater	Streams.
Taxon specificity	Generic	Macroinvertebrates, periphyton.
Toxicant specificity	Generic	Pesticides, organic pollution.
Application	Known in science	
Public availability	Source code R package	https://cran.r-project.org/web/packages/streambugs/index.html

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	Biomass pools for macroinvertebrate taxa, periphyton, fine and suspended particulate organic matter.
Endpoints	Population structure Biomass	
Space	Implicit	
Time		

Criteria	Categories	Comments
Exposure / effects	Varying concentrations Pulse exposure	Direct effects modelled as increase in mortality. The increase in mortality depends on classification of a taxon as sensitive or insensitive according to the SPEAR data base; only sensitive taxa are affected depending on the toxicant concentration in the water. In a later application (Kattwinkel et al. 2016), a species-specific linear increase in acute mortality with log(concentration in water) is modelled starting at an assumed NOEC of 0.5 x log(LC50) of a species.
Abiotic environment	Food limitation Temperature Light Others	Current speed, substrate.
Biotic environment	Intraspec. competition Interspec. competition Predation	
Individuals	None	
Populations	Biomass	
Calibration	Field data Laboratory data	Mass balance for different species calibrated by scaling according to the rules of the metabolic theory of ecology (MTE). Parameterization of feeding and habitat preferences from CASiMiR and www.freshwaterecology.info data bases. Sensitivity based on SPEAR and saprobic data base. Additional calibration by adapting food web parameters until the model predictions match better the occurrence / absence patterns of species in 87 samples from 4 sites ("learning from data", Bayesian inference).
Programming language	-	R

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data Field data	Initial population size does not affect equilibria after some run-time. Calibration using multiple independent data ("soft validation"). The model was able to reproduce the community composition in those data without re-calibration to each specific data set. Predictions improved when pesticides were considered, showing their significance.
Sensitivity analysis	No	
Uncertainty analysis	Yes	Monte-Carlo simulation. Endpoint for the analysis was the probability of predicting considerable population densities (levels that are detectable in the field). The results were quite stable. Different types of functional response did not affect the output.
Documentation	Scientific publication	Difficult to read, but comprehensive.

Assessment

Criteria	Description
Strengths	Comparably low number of free parameters due to the application of rules for the parameterization based on MTE and ecological stochiometry. Sophisticated mass balancing.
Theoretical uncertainties	
Empirical uncertainties	
Parametric uncertainties	
Temporal uncertainties	Model does not capture short-term population dynamics but predicts a stable state.
Conclusions	

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Schuwirth and Reichert (2013)	Freshwater organisms	Generic	Original publication.
https://www.eawag.ch/de/abteilung/siam/projekte/streambugs/			
Model applications			
Kattwinkel et al. (2016)	Macroinvertebrates	Thiacloprid	Application to data from a mesocosm study. Direct effects were modelled to linearly increase mortality with logConcentration if a certain threshold (0.5 log-units below LC50) was exceeded. 85 % of observed data points were within the 95 % CI, but no validation with independent data. Emergence process and sublethal effects turned out to be potentially relevant for future extensions.

1.5.1.5 **Chemostat Model with DEB (Kooi et al. 2008)**

Chemostat models describe the dynamics of several populations in a homogenous environment (a chemostat), and explicitly consider nutrients. Here we review an example of Kooi et al. (2008b). Populations are treated unstructured: a single ODE for the biomass over time of each population (Marr-Pirt model). Populations interact by feeding on each other, or by competing for the same food source. Toxicants are accumulated from water and food, and the internal concentration affects a physiological process in the population (maintenance, assimilation or mortality). This approach to toxic effects is directly comparable to that used in DEB-based models.

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	The unstructured models deal with the food chain/community level. These models are rather simplistic, and could be used at lower tiers.
Model purpose	Scientific	The primary objective seems to be scientific: To study how chemical stress changes the dynamics of several populations interacting with each other in a chemostat environment.
Questions / processes	Effect propagation Population recovery	The model follows the dynamics of several population as they interact in a homogeneous environment (chemostat). The populations are unstructured, which is most suitable for small (single-celled) organisms.
Environmental domain	Freshwater	In principle, the models are generic, although the assumption of a homogeneous environment, with a constant inflow of medium, is most relevant for aquatic systems. The unstructured treatment of populations makes these models most suitable for single-celled organisms.
Taxon specificity	Generic	
Toxicant specificity	Generic	These models are generic.
Application	Established in science	Ecosystem models with unstructured populations are well established in ecology, but do not seem to have been applied much to toxicant stress, and certainly not in a regulatory setting.
Public availability	-	There seem to be no publicly available versions of these models.

Variables and Parameters

Criteria	Categories	Comments
Entities	Body burden Population biomass	There are equations for the biomass of each population, and for the internal concentration in each population. The populations are unstructured, so there is no distinction between individual and population biomass.

Criteria	Categories	Comments
Endpoints	Population biomass Body burden	Endpoints are on the population dynamics. Not only the size of each population, but also its dynamic behaviour over time. For example, under certain conditions, population biomass may start to show cyclic or even chaotic behaviour. Toxic stress changes the positions where these changes in model behaviour occur.
Space	No spatial context	The environment is taken homogeneous, with a constant inflow of nutrients and a constant outflow of nutrient and biomass (chemostat).
Time	-	Generally, these models focus on the long-term behaviour of the system.
Exposure / effects	Varying concentrations Chronic vs. pulse	In the publications examined, the inflow of the toxicant into the system is taken as constant. However, this inflow can easily be taken as a function of time. Lethal or sublethal effects are calculated from the internal concentration using a DEB model.
Abiotic environment	Food limitation	Food (or better: nutrients) is explicitly followed. Nutrients constantly flow into the chemostat, are mixed and used by the populations, and flow out. When the populations reach a considerable size, food limitation will occur.
Biotic environment	Intraspecific competition Interspecific competition Predation	Populations may compete for the same food source or prey on each other. No other interactions considered but through feeding relationships.
Individuals	None	Individuals are not considered; the population is modelled as a single super-organism.
Populations	Biomass	One ODE for each population, following only its biomass as the single state variable.
Calibration	Laboratory data	The authors refer to other studies for the sources of their parameters, which are likely based on lab experiments with single-celled species. Bontje et al. provide examples were models from this category were fitted to experimental data for population biomass over time.
Programming language	-	Specialist software used for the bifurcation analyses.

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	No validation attempts. Bontje et al shows that similar models could be fitted to experimental data.
Sensitivity analysis	Yes	The bifurcation analysis represents an advanced form of sensitivity analysis. Here, only the dilution rate and nutrient concentration are changed to see the effect on the system dynamics.
Uncertainty analysis	No	

Criteria	Categories	Comments
Documentation	Scientific publication	The models are explained in detail in several publications. No structured documentation (e.g., TRACE) available.

Assessment

Criteria	Description
Strengths	The simplicity of the model allows for a structured analysis of long-term system dynamics (through bifurcation analysis). Further, the effects of several feedbacks can be analysed (e.g., via the food source and via the concentration of toxicants in the system). Toxicant stress is included as a relationship between internal concentrations and physiological processes, which is more mechanistic than what is used in other models.
Theoretical uncertainties	The simplicity of the model also means that the modelled environment is hardly realistic. The behaviour of several unstructured populations in a chemostat environment is difficult to extrapolate to the field situation.
Empirical uncertainties	
Parametric uncertainties	The parameterisation of the populations may affect the results of the system. These parameters may not be simple to derive for specific species.
Temporal uncertainties	Constant inflow of toxicant and nutrient is assumed.
Conclusions	These models provide insight into the effects of toxicants on the long-term dynamics of simple systems. They could be used to study the effects of transient toxicant stress. However, such a detailed analysis of single-celled populations (in an unrealistic environment) may not be a primary concern for the risk assessment of PPPs.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kooi et al. (2008b)	Unicellulars	Generic	
Kooi (2003)	Unicellulars	None	Discussion of a range of unstructured ecosystem modelling in a chemostat setting, and how bifurcation analysis can be used to interpret them.
Model applications			
Kooi et al. (2008a)	Unicellulars	Generic	Extension of the model with nutrient cycling (via two detritus pools), an additional prey (there are now two consumers of nutrients), and a top predator.
Bontje et al. (2009a)	Algae (*Cryptomonas*)	Prometryn, parathion	Focus on one population (algae), with nutrient recycling in a closed system. Model is fitted to experimental data.

Citations	Taxa	Chemicals	Comments
Bontje et al. (2009b)	Algae (*Cryptomonas*) and three species of ciliates	None	Fitting a related unstructured model for data on prey (algae) and predator (ciliate) biomass over time. Model includes a detritus pool.
Bontje et al. (2011)	Unicellulars	Prometryn	Extension of the model with a sediment compartment, nutrient recycling (three detritus pools), a population of consumers in the sediment, and a top predator.

1.5.2 Saltwater Models

1.5.2.1 ECOWIN (Ferreira 1995)

The model of Ferreira (1995) simulates a simplified estuary ecosystem and is part of the ECASA project that provides information for the establishment of aquacultures. The community consists of biomass pools for phytoplankton and different weight classes of oysters. Mass balance for nitrogen with a simplified nitrogen cycle. Transport of nutrients, suspended matter and phytoplankton between adjacent compartments (boxes) through river flow is explicitly modelled. Resuspension of nutrients from sediment into the water through turbulence considered. Toxicants are not considered in the applications of the model. The model has been applied to some case studies of estuaries, the first one in Ireland was used for assessment.

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	
Model purpose	Scientific	
Questions / processes	Effect propagation Others	Model the effect of various stressors on estuarine ecosystems for the aquaculture industry. Demonstrate the application of an object-oriented approach in ecosystem modelling.
Environmental domain	Freshwater	Estuaries
Taxon specificity	Specific	Oysters and phytoplankton.
Toxicant specificity	-	No toxicants
Application	Established in science	
Public availability	Software extension	Can be requested at https://www.long-line.co.uk/site/products/aquaculture/ecowin/

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	Suspended particulate matter, oyster and phytoplankton biomass.
Endpoints	Biomass	
Space	Boxes	Boxes represent adjacent coastal stretches at the landscape scale (km²).
Time	Hours	Time step of two hours.
Exposure / effects	No exposure	No applications of toxicant mixtures, though modelling of mixtures with additive effects might be possible. Only acute effects.
Abiotic environment	Food limitation Temperature Light Water quality	Sediment transportation, advection (river flow), salinity.

Criteria	Categories	Comments
Biotic environment	Intraspecific competition Predation	Anthropogenic seeding and harvesting of oysters is simulated.
Individuals	None	
Populations	Biomass	
Calibration	Field data	Parameters for transportation of sediment between boxes fitted from bathymetric data applying an external hydrodynamic model.
Programming language	C++, Turbo Pascal	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	No independent data	Validation with data used for calibration, as the application was site-specific. Phytoplankton biomass was slightly underpredicted, suspended matter well predicted, growth of oyster biomass was slightly overpredicted but reasonable.
Sensitivity analysis	Yes	
Uncertainty analysis	No	
Documentation	Scientific publication Website	Description very much focused on programming issues, difficult to understand the conceptual model.

Assessment

Criteria	Description
Strengths	Relatively low complexity, reasonable accordance of model predictions with observations from the data set used for calibration.
Theoretical uncertainties	Highly simplified food web. No non-human predators and competitors of oysters considered, no interactions of phytoplankton with different functional groups.
Empirical uncertainties	Many parameters such as resuspension of organic material through turbulence can be parameterized only with high uncertainty due to high spatial and temporal variation.
Parametric uncertainties	All phytoplankton species clumped together in a common biomass pool.
Temporal uncertainties	The model assumes constant tidal current cycles, but current can vary due to different water levels of the stream.
Conclusions	ECOWIN is a highly simplified simulation of estuarine ecosystems with a moderate level of complexity. The model lacks a toxicant module but was able to reproduce population dynamics in oyster farms reasonably well.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Ferreira (1995)	Oysters, phytoplankton	None	Original publication of the model.
Sequeira and Ferreira (2005)	Oysters, phytoplankton	None	ECASA model description
http://www.ecowin.org/			Old webpage, seems not be maintained anymore.
https://www.longline.co.uk/site/products/aquaculture/ecowin/			Modern webpage.
Model applications			
Ferreira et al. (1998)	Oysters, phytoplankton	None	First application to assess carrying capacity of Irish oyster banks.

1.5.2.2 **NEMURO (Kishi et al. 2007)**

NEMURO (North Pacific Ecosystem Model for Understanding Regional Oceanography) was developed by Kishi et al. (2007) as a consensus prototype lower trophic level ecosystem model for PICES CCCC (North Pacific Marine Science Organization, Climate Change and Carrying Capacity program). This conceptual biomass model represents the minimum trophic structure and biological relationships between and among all the marine ecosystem components thought to be essential in describing ecosystem dynamics in the North Pacific. NEMURO contains > 70 parameters and is thus less complex than CATS, CASM and AQUATOX. Cycling of the limiting factors N and Si in the water, and the functional groups phytoplankton, small, large and predatory zooplankton are explicitly simulated. Seasonally, large zooplankton enters and leaves the simulation to consider vertical migration from lower water layers outside the modelled regions. Gelatinous zooplankton represents the top predator in the model and is considered to include the biomass of all higher trophic levels not explicitly simulated. However, the model can be routinely coupled to an age class-structured bioenergetics model for the fish predators Saury and Herring to get a full food web-model (NEMURO.FISH).

General Properties

Criteria	Categories	Comments
Biological level	Ecosystem	
Model purpose	Scientific / Regulatory	
Questions / processes	Population recovery	Analyse effects of climate change on structure and function of oceanic ecosystems.
Environmental domain	Marine	Top water layer in the Northern Pacific.
Taxon specificity	Generic	Phytoplankton, small, large and predatory zooplankton, (fish).
Toxicant specificity	-	No toxicants
Application	Established in science	
Public availability	Software extension	Executable box models in FORTRAN, MATLAB and others. Source code is freely available.

Variables and Parameters

Criteria	Categories	Comments
Entities	Biomass Others	11 state variables: nitrate, ammonium, small and large phytoplankton biomass, small, large and predatory zooplankton biomass, particulate and dissolved organic nitrogen, particulate silica, silicic acid concentration, fish (in an extension).
Endpoints	Biomass Water quality	Biomass of each functional group and amount of N and Si.
Space	No spatial context	The model describes the average conditions in a water column of about 1 m² width of the upper mixed layer of an ocean.
Time	Days	The model is typically run for 5 - 10 years before reaching a stable state that exhibits expected dynamics of the state variables.
Exposure /	No exposure	

Criteria	Categories	Comments
effects		
Abiotic environment	Food limitation Temperature Light Water quality	
Biotic environment	Intraspecific competition Interspecific competition Predation	
Individuals	None	
Populations	Biomass	
Calibration	Field data Laboratory data	Multiple applications with specific calibrations, particularly on plankton. Original calibration from two high sea stations in the western and eastern part of the Northern Pacific.
Programming language	FORTRAN, MATLAB and others.	

Evaluation and Documentation

Criteria	Categories	Comments
Validation	Field data	Reasonable reproduction of seasonal patterns in the dominance of different functional groups of marine plankton.
Sensitivity analysis	Yes	Monte Carlo simulations. 8 particularly important parameters in NEMURO.
Uncertainty analysis	Yes	Predictions of nutrient fluxes by NEMURO were satisfactorily when compared to predictions of other marine models.
Documentation	Scientific publication	Well-structured and comprehensive documentation.

Assessment

Criteria	Description
Strengths	Low complexity but comparably high ecological realism. Active use and development by a large number of experts. Reasonable results after comprehensive testing with independent data justifies confidence in the model predictions.
Theoretical uncertainties	No anthropogenic effects included yet (toxicants, fishery etc.). Incomplete food web (missing top predators, decomposers) may underestimate effects of ecological disturbance.
Empirical uncertainties	Food preferences of zooplankton are difficult to quantify and therefore subject to guesses.
Parametric uncertainties	Early life stages of fish (eggs and larvae) ignored.
Temporal uncertainties	No differentiation between conditions at day / night.

Criteria	Description
Conclusions	Successful compromise of minimal complexity and maximal ecological realism. The biggest advantage is the large number of applications / testing. Could be valuable for risk assessment in estuaries, if a reasonable toxicity module is integrated.

Publications

Citations	Taxa	Chemicals	Comments
Model description			
Kishi et al. (2007)	Marine phytoplankton, zooplankton	None	Original publication of NEMURO.
https://www.pices.int/members/task_teams/Disbanded_task_teams/MODEL.aspx			Website
Model applications			
Kishi et al. (2011)	Marine phytoplankton, zooplankton, fish	None	Review of NEMURO and NEMURO.FISH applications. Various extensions such as carbon cycle, microbial food web, additional fish species, three-dimensional space, and advection of zooplankton at coastal regions. Generally reasonable accordance of predictions and test data, but sometimes underpredictions of zooplankton biomass.
Fiechter et al. (2015)	Marine phytoplankton, zooplankton, chinook salmon	None	Good prediction of observed growth rates of salmon during good and bad years between 1984 and 2006.

2 Evaluation of Effect Models for the Risk Assessment of Pesticides

2.1 Introduction

In the second part of this report, we reviewed 10 selected models in detail. Most important criteria for the selection of models were the potential for use in the risk assessment of pesticides, i. e. the potential applicability in the scheme of regulatory risk assessment and the developmental state. Accordingly, we preferred models that have been already proposed in dossiers for the registration of active substances or plant protection products. Additionally, we aimed at covering models that differ in their spatial context and that address all levels of biological organization (individual, population, community and ecosystem).

The evaluation of the spatially explicit population model ALMaSS addressed an application for small mammals and covered also a separate population model for the wood mouse from Liu et al. (2013) that may be applied in a similar way. For the community level, by the time of model selection in 2017, no simulation model has been identified that was considered potentially fit for application in risk assessment. Instead, $SPEAR_{pesticides}$ was evaluated that has been developed as an indicator system for the assessment of pesticide exposure based on observed environmental effects. However, the SPEAR approach may be applied also in the opposite way for the assessment of effects based on observed or predicted pesticide concentrations. Tab. 5 provides an overview of the selected models.

Table 5: Effect Models Evaluated in Detail

Model name	Organization level	Organism group	Spatial context	Most relevant citations
GUTS	Individual	Generic	None	Jager et al. (2011), Jager and Ashauer (2018b)
DEBtox	Individual	Generic	None	Jager and Zimmer (2012), Jager (2019)
IDamP	Population	Freshwater invertebrates	None	Preuss et al. (2009a)
IBM *Chaoborus* population model	Population	Freshwater invertebrates	Metapopulation	Strauss et al. (2016), Strauss (2017)
MASTEP	Population	Freshwater invertebrates	Spatially explicit	Van den Brink and Baveco (2009)
SpringSim	Population	Soil organisms	Spatially explicit	Meli et al. (2013)
eVole	Population	Small mammals	Spatially explicit	Wang (2013), RIFCON (2018)
ALMaSS + Woodmouse Model	Population	Small mammals (in ALMaSS also birds, non-target arthropods)	Spatially explicit	Topping et al. (2003), Liu et al. (2013)
$SPEAR_{pesticides}$	Community	Freshwater invertebrates	None	Liess and von der Ohe (2005)
AQUATOX	Ecosystem	Freshwater organisms	Metapopulation	Park et al. (2008), Park and Clough (2018)

For the model description and evaluation, we used information that has been made publicly available to the scientific community. This included articles in peer-reviewed scientific journals, publicly available model documentations, and model demonstrations found on web pages and posters. Additionally, we considered non-publicly available documentations of more recent model versions if proved by the UBA. These documentations were typically supplied as supporting material for modelling reports that have been proposed to authorities of the EU Member States for the regulatory risk assessment of pesticides. Permission for use was obtained from the authors prior to the publication of this report.

The description and evaluation of each model has been organized in four sections: First, the general information provides an overview in continuous text form on the background and concept of a model, and on the current status in terms of development and applications. Second, a detailed model description is provided that has been structured according to Tab. 1 in the EFSA Sci. Op. on GMP (2014b)[1]. This documentation scheme was developed as a template for a summary document that shall be provided to risk assessors along with a modelling study. It includes a comprehensive list of specific questions and topics to be addressed, and covers important aspects of the modelling cycle such as the problem definition, the supporting data, the model concept, the formalization, the software implementation, the parameterization, a sensitivity analysis, and the model validation (comparison of model output with observed data).

However, the documentation scheme addresses both information on a model in general and on a specific model application for a given pesticide use (the "regulatory model"). While some aspects of the environmental scenario and the parameterization may be considered for the model in general, other aspects are case-specific and need to be described and evaluated separately for each model application. E. g., functions that relate environmental conditions such as day length and temperature to latitude, or the parameterization for the physiology of a model species may be considered built-in into the general model; they are not expected to change as long as the model is used within its domain of applicability. In contrast, the setting and parameterization of a particular environmental scenario including food supply and landscape composition is case-specific and needs to be documented and evaluated for each regulatory model. Therefore, not all questions could be addressed in part 2 of this report. To address those questions that relate to specific regulatory models (the environmental scenario and its parameterization, sensitivity and uncertainty analyses, and the model use for risk assessment), we used information on publicly available case studies for model demonstration. Typically, the models were presented to the scientific community using one or several default scenarios and parameterization that may be considered representative for their potential use in risk assessment. However, the aim of these case studies was the demonstration of the general model applicability and not the risk assessment of a specific pesticide; therefore, we left out the last part of the documentation scheme that deals with case-specific conclusions for the regulatory risk assessment.

Third, each model description is followed by a structured evaluation of the potential for risk assessment from a scientific point of view. This evaluation was based on a checklist provided in Appendix B of the EFSA Sci. Op. on GMP (2014b). This checklist was developed for risk assessors to conduct a comprehensive evaluation of a model and of its application for the risk assessment of a pesticide. A summary and conclusions of the evaluation can be found at the end of that section. Again, case-specific questions on a regulatory model could not be addressed for the model in general and some questions regarding a specific risk assessment were excluded, but we considered case studies published for model demonstration.

[1] EFSA PPR (2014): Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA journal 12(3): 3589.

Fourth, the evaluation was supplemented by a qualitative assessment of uncertainties, listing sources for the potential over- and underestimation of real risks when applying the model for the environmental risk assessment of pesticides. This list was inspired by Appendix C in the EFSA Sci. Op. on GMP (2014b) that provides criteria for a qualitative assessment of uncertainty in ecological modelling.

2.2 GUTS

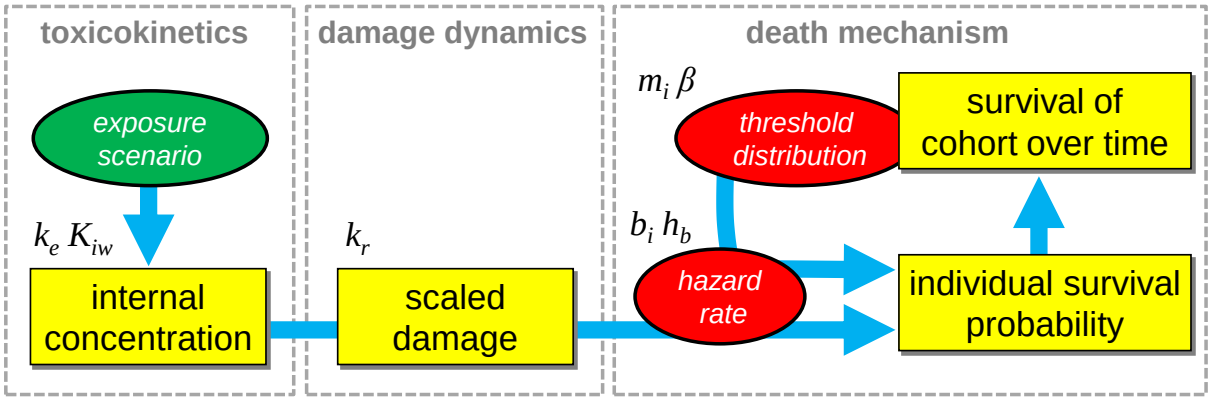
Evaluation by Tjalling Jager

2.2.1 General Information

2.2.1.1 Background and Concept

GUTS stands for the General Unified Threshold model for Survival. It is not a single model, but rather a framework from which specific toxicokinetic-toxicodynamic (TKTD) models for the endpoint survival can be derived as special cases. Survival modelling has a long history in ecotoxicology (at least half a century), which has led to a variety of, seemingly very different, models. GUTS unifies all these previous models into a consistent over-arching framework. It must be stressed that GUTS only deals with effects on survival, and potentially other all-or-nothing endpoints such as immobility. GUTS has been presented in a paper in the open literature (Jager et al. 2011)[2], and an extensive e-book has more recently been released (Jager and Ashauer 2018b)[3] as part of a Cefic-LRI-funded project (see http://www.debtox.info/book_guts.html). The model description below is largely taken from that book. The full structure of GUTS is schematically shown in Figure 1. In practice, simplified models will be used that are special cases of this framework.

Figure 1: GUTS – Schematic Representation



Symbols represent model parameters for the various processes: k_e (elimination rate constant), K_{iw} (bioconcentration factor), k_r (damage repair rate), m_i (median of threshold distribution for effects), β (width of threshold distribution), b_i (killing rate; how fast the hazard rate increases above the threshold) and h_b (background hazard rate). Graph reproduced from Jager and Ashauer (2018b).

A chemical first needs to be taken up from the environment before it can exert a toxic effect. Hence, the first module is a toxicokinetics (TK) model, and GUTS applies the simplest version: the one-compartment model with first-order kinetics (see assumptions 1 and 2 below). If needed, more complex TK models may be inserted. The toxicodynamics (TD) part is made up of two modules: damage dynamics and the death mechanism. The internal concentration leads to damage, which is repaired at a certain rate (assumption 3). Damage causes mortality (assumption 4), using two mechanisms for death: each individual has a probability to die, which is increased by the damage above a threshold, and each individual has a different value for the threshold, drawn from a log-logistic frequency distribution (assumption 5). Thus, the complete set of assumptions underlying GUTS is:

[2] Jager, T., C. Albert, T. G. Preuss and R. Ashauer (2011): General Unified Threshold Model of Survival - a Toxicokinetic-Toxicodynamic Framework for Ecotoxicology. Environmental Science & Technology 45(7): 2529-2540.

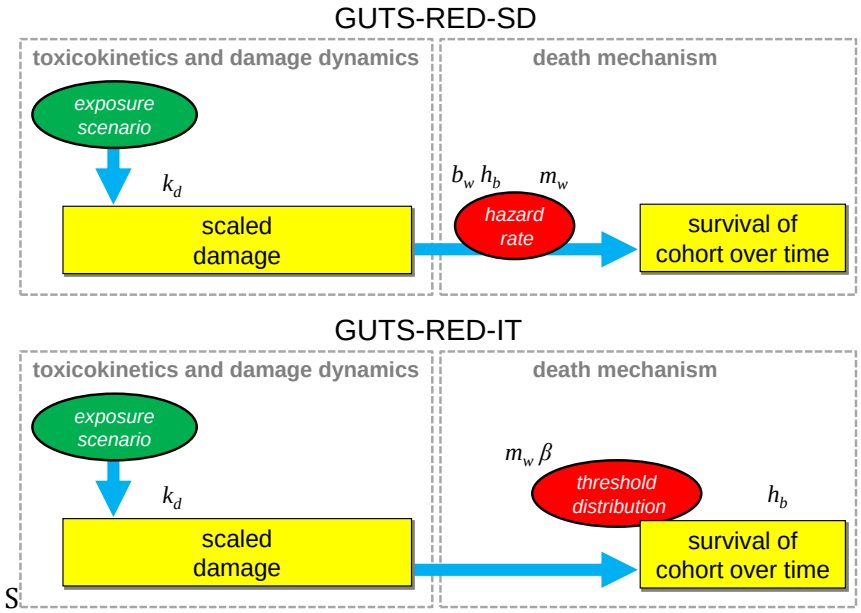
[3] Jager, T. and R. Ashauer (2018b). Modelling survival under chemical stress. A comprehensive guide to the GUTS framework. Version 2.0, 8 December 2018, Toxicodynamics Ltd., York, UK. Available from Leanpub, https://leanpub.com/guts_book.

1. Chemicals first need to be taken up into the body of the organism before they can exert their effect.
2. The organism is treated as one homogeneous (well-mixed) compartment. The uptake flux from the environment into the organism is proportional to the external concentration, and the elimination flux from the organism to the environment is proportional to the internal concentration. The external concentration is not influenced by uptake into the organism (i.e., the environment is infinitely large and well mixed).
3. The accrual flux of damage is proportional to the internal concentration, and the repair flux is proportional to the damage level. The internal concentration is not influenced by damage accrual in the organism, and damage is treated as one homogeneous (well-mixed) compartment.
4. The toxic effect (the death mechanism) is driven by the damage level.
5. Each individual organism has a threshold for the damage level. The value for this threshold differs between individuals and can be described by a log-logistic distribution. When the damage level is below the threshold, there is no effect of the chemical on mortality. When damage exceeds the threshold value, the hazard rate due to the chemical stress becomes proportional to the value of the damage level above the threshold.
6. Background mortality is independent of the mortality caused by the toxicant. For short toxicity tests, the background hazard rate can be taken constant (this represents deaths due to accidents, and not due to ageing).
7. The organism does not change over time. In other words: the model parameters remain constant.

Assumption 1 is the basic tenet of TKTD modelling, and of ecotoxicology in general: effects cannot be understood from external concentrations; the chain of events starts with internal concentrations (Escher and Hermens 2002, Escher et al. 2011). Point 2 specifies the set of assumptions that leads to the classic one-compartment TK model. This model has a long history in science, as well as in regulatory settings (OECD 2012), and often provides a good explanation of body residues over time. It is the simplest possible TK model, and owing to the general paucity of data in ecotoxicology, usually the only one that can be used in TKTD modelling. Assumptions 3 and 4 deal with the damage module. For many chemicals, it is not the total concentration of the parent compound in the body that directly drives the toxic effect; it might be a metabolite, or reactive damage to macromolecules, or the disruption of acetylcholinesterase, etc. To decouple the effect dynamics to some extent from the parent compound's kinetics, a simple one-compartment, first-order, damage module is added (as first proposed by Ankley et al. 1995). In the future, more complex models may be inserted here for specific cases (e.g., based on explicit receptor kinetics or adverse-outcome pathways), but due to the general lack of specific data, the one-compartment damage module is the starting point. Assumption 5 deals with the death mechanism; it must explain why organisms die, and why they do not all die at the same time under the same conditions. In reality, there will be many factors that play a role in determining whether an individual dies or not, but these are generally condensed into a chance process. The set of assumptions under point 5 combines the two classic models for survival: stochastic death (SD, all individuals are the same, death is a chance process at the level of the individual) and individual tolerance (IT, individuals differ in sensitivity, death is deterministic for the individual but a chance process at the level of a cohort). These two mechanisms have a long history in ecotoxicology under various names: the models focussing on IT, for example, as critical-body residue (CBR) models (e.g., Mackay et al. 1992), and those focussing on SD as the survival model of the DEBtox software (e.g., Bedaux and Kooijman 1994). These two options seem to be the only ones that have been used in ecotoxicology so far to cover the death mechanism. Assumption 6 deals with the background mortality, which is treated as a constant chance of accidental deaths (although more complex ageing modules can be inserted). Assumption 7 is a practical one: if the organism changes considerably over the duration of an experiment, it will be impossible to fit the model, unless extensive data are available to identify and quantify these changes. This final assumption is especially critical when GUTS is to be used to extrapolate from short acute toxicity tests to much longer field scenarios. The full GUTS model, as depicted in Figure 1, has 7 parameters, which is too much to fit using only standard ecotoxicity test data for survival, such as the 4-day acute

fish test. Furthermore, there is usually no information on body residues or toxicokinetics available to allow estimation of all parameters of the TK and damage modules. Therefore, GUTS is generally applied in a reduced form, especially as the two special cases GUTS-RED-SD and GUTS-RED-IT, schematically shown in Figure 2. These two special cases have a history in ecotoxicology, e.g., as the 'DEBtox survival model' and 'dynamic CBR model'. Both models have just four parameters, which can be estimated from survival data over time, as produced by standard acute toxicity tests (an example fit is provided in the next section).

Figure 2: GUTS – The Two Reduced Cases of GUTS Most Commonly Applied in Ecotoxicology



Schematic representation of the two reduced cases of GUTS that are most commonly applied in ecotoxicology. Symbols represent model parameters (see Table 6). The new model parameter k_d replaces the two rate constants (k_e and k_r) of the full model. It is referred to as the 'dominant rate constant', and is the one-compartment approximation of the two-compartment system of toxicokinetics and damage. Graph provided by Tjalling Jager.

In the reduced models, the TK and damage modules are collapsed into a single one-compartment model for scaled damage. The resulting rate constant (k_d) is referred to as the 'dominant' rate constant. This rate constant represents the combination of the two initial processes in the model: toxicokinetic uptake/elimination and damage repair (the slowest process will dominate the overall kinetics of damage). Each model in Figure 2 focusses on a single death mechanism, either stochastic death (SD) or individual tolerance (IT). It turns out to be very difficult to select one of these two mechanism as the most realistic one, and in reality, it is likely that both play a role (the full GUTS model combines both mechanisms, but standard toxicity tests do not provide sufficient detail to fit the full model). Using these two extreme views to analyse the data, and to make predictions, should provide a good coverage of reality (see Ashauer et al. 2013), and can be seen as a form of structural sensitivity analysis. For ERA purposes, one could thus decide to use both models and focus on the most conservative result; this procedure was recently adopted by EFSA in the scientific opinion on TKTD models (EFSA PPR 2018).

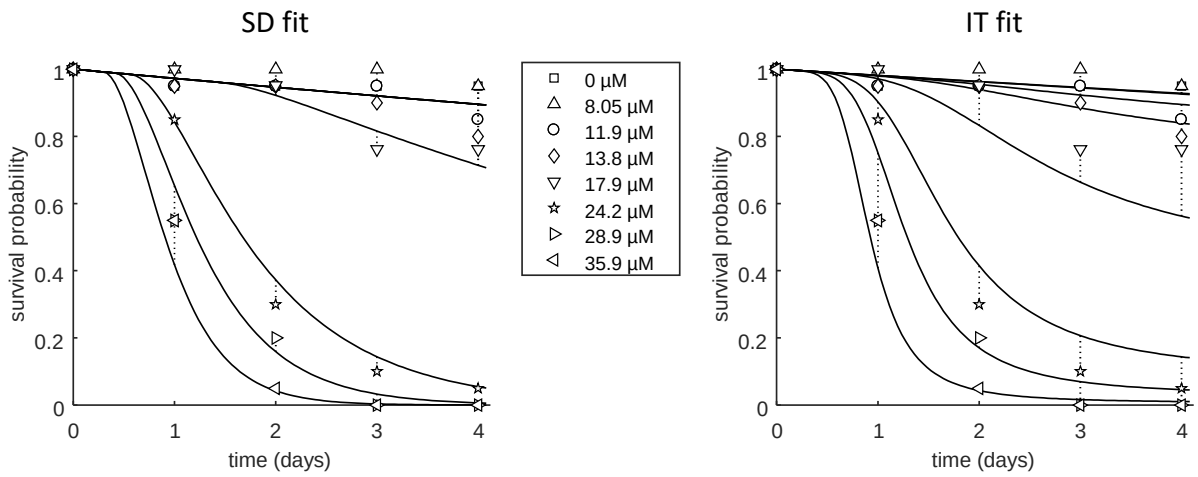
The reduced GUTS models can be easily fitted to acute toxicity data following from many standard test protocols (although meaningful model application requires observations at multiple time points). The calibrated model can subsequently be used to make predictions for other exposure scenarios (e.g., constant to time-varying exposure, and short to long exposure). Such extrapolations rest on the assumption that the model is true (that the mechanisms modelled are a good representation of reality), and that the parameters established in the calibration remain constant and are relevant for the new expo-

sure scenario as well. TK is influenced by body growth (dilution by growth and changes in surface:volume ratio) and reproduction (transfer of chemicals to offspring), and generally affected by environmental temperature. The intrinsic sensitivity of the organisms (e.g., reflected in the threshold for effects) may also depend on temperature, and possibly also on the presence of other stresses such as starvation and disease. At this moment, the effect of these factors will be difficult to predict in general, and extrapolations (far) beyond the conditions of the test will thus be accompanied by additional uncertainty.

Example Fit

As an example, Figure 3 shows the fit of the two reduced models on a data set for propiconazole in the amphipod *Gammarus pulex* (Nyman et al. 2012). Both reduced models provide, visually, a good fit to the data. The fit of the IT model is, however, poorer, and based on the Akaike information criterion, the SD model provides a better explanation of this data set (ΔAIC = 7.8). However, the IT model better captures the small dose-related effects at the lowest exposure concentrations. The two models yield a similar estimate of the LC50-4d (Table 6), but can lead to different predictions when extrapolating beyond the data set used for calibration. These fits were performed with the BYOM platform and GUTS package for Matlab (http://www.debtox.info/byom.html).

Figure 3: GUTS – Example Fit of the Two Reduced Models



Example fit of the two reduced models shown in Figure 2 on a data set for propiconazole in *Gammarus pulex* (Nyman et al. 2012). Parameter estimates shown in Table 6. Fits performed with BYOM. Concentrations in the legend are mean measured concentrations (measured daily, all concentrations within 15% difference). Graphs provided by Tjalling Jager.

This example demonstrates how GUTS can be used to analyse toxicity data and derive useful parameters, such as a (median) threshold for effects (m_w), as well as more familiar output, such as the LC50 after 4 days constant exposure, or in general an LCx,t for any effect level x and exposure duration t. In this case, the 4-d LC50 from the GUTS fit is very similar to the value given in the publication resulting from fitting a dose-response curve: 19.2 (17.6–20.9) µM. The parameters, with their joint confidence interval, could subsequently be used to predict survival probability due to another exposure profile (e.g., the output from a fate model), as demonstrated in several studies (Nyman et al. 2012, Ashauer et al. 2013, Ashauer et al. 2016). This case study was also used in the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) to demonstrate the proposed workflow for the application of GUTS in ERA: the model is fitted to data for constant exposure, validated with additional studies on pulsed exposure, before it can be used for extrapolation to FOCUS scenarios.

GUTS could, in principle, also be used to aid various extrapolations, other than between exposure scenarios. Some work has been done to study the possibilities for extrapolation between life stages (Gerritsen et al. 1998, Jager et al. 2016a), species (Baas and Kooijman 2015) and between chemicals (Jager and Kooijman 2009, Ashauer et al. 2015). These studies have shown the potential for GUTS in these areas, but require further research before they can be used routinely in ERA for predicting model parameters for new chemicals or species. Also for mixture toxicity, several studies have shown that GUTS can be used to analyse data for combined effects (Baas et al. 2007, Ashauer et al. 2017). It is also possible to predict the effects of untested mixtures from the model parameters for single components, but such predictions will have to be based on the assumption that the components do not interact (interactions cannot yet be predicted).

Table 6: GUTS – Example Parameter Estimates

Symbol	Parameter	SD fit (95% CI)	IT fit (95% CI)	Unit
k_d	Dominant rate constant	2.2 (1.6-3.3)	0.75 (0.56-0.98)	d^{-1}
m_w	Median of threshold distribution	17 (16-18)	18 (15-21)	µM
b_w	Killing rate	0.13 (0.086-0.20)	Not used in IT	$\mu M^{-1}\ d^{-1}$
β	Width of threshold distribution	Not used in SD	7.1 (5.2-9.3)	[-]
h_b	Background hazard rate	0.028 (0.013-0.050)	0.019 (0.0050-0.041)	d^{-1}
Derived from model parameters				
LC50 4d	Concentration associated with 50% mortality, after 4-day constant exposure	19 (18-22)	19 (16-22)	µM

Parameter estimates for the fit in Fig. 3 with 95% likelihood-based conf. intervals. Fits performed with BYOM. The LC50 is not a model parameter but a model prediction that follows from the model parameters (and their joint uncertainty). Data from Nyman et al. (2012, edited).

2.2.1.2 Status of the Model

A predecessor of GUTS, the DEBtox-survival model (currently viewed as one of the special cases of GUTS), has been included in OECD/ISO guidance on the statistical treatment of ecotoxicity data (ISO 2006, OECD 2006) under the header 'biology-based methods'. This work is also mentioned in REACH guidance (ECHA 2008). GUTS analyses have been submitted as part of dossiers for risk assessment of PPPs. Relevant examples of potential use in risk assessment have been published in the open literature (Ashauer et al. 2013, Ducrot et al. 2015). GUTS is one of the models treated in detail in the EFSA scientific opinion on TKTD modelling for aquatic ERA of PPPs that concluded that GUTS is "ready for use in aquatic ERA".

It is important to stress that GUTS could potentially be used in ERA in different ways. The most prominent ones are:

1. Analysis of data from toxicity tests, using all of the data from the test over time. The median threshold from GUTS (m_w) can be used as summary statistic (for SD, it is a true no-effect concentration), or GUTS can be used to calculate LCx,t, for any effect level x and any exposure time t. Furthermore, the analysis can reveal inconsistencies in the experimental data, or between data sets, and thus be used as a quality check on the data. Additional advantage over descriptive methods (such as fitting dose-response curves) is that data from non-standard test designs could be used for calibration (and calculation of LCx,t) without problem (e.g., when the exposure concentration has not been constant over the test duration).
2. Extrapolation of effects to untested environmental conditions. The calibrated model can be used to predict survival over time for a different exposure scenario (Nyman et al. 2012, Ashauer et al.

2013, Ashauer et al. 2016). For example, the model may be calibrated on standard toxicity data (e.g., a 4-day test at constant exposure) and used to predict long-term effects on survival due to an exposure profile from a fate model.

3. As individual-level effects module in population/community models (see e.g., Gabsi et al. 2014c, Dohmen et al. 2016). GUTS thus can be (and has already been) implemented in higher-level models (although it is good to consider the uncertainties of the model in light of this application; see end of this evaluation).
4. Extrapolation between chemicals, species and life stages. Several proofs-of-concept have been published (Gerritsen et al. 1998, Jager and Kooijman 2009, Ashauer et al. 2015, Baas and Kooijman 2015, Jager et al. 2016a).

Application 1 has been most common so far, and also the one that the OECD/ISO guidance (ISO 2006, OECD 2006) focusses on. For ERA of PPPs, application 2 (and linked to that, application 3) is of specific interest. For application 4, more structured research will be needed to fulfil this purpose in ERA. From the list of potential applications, the recent EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on application 2 and proposes a specific workflow to that end: 1) calibrate the two reduced GUTS models (SD and IT) to toxicity data for constant exposure, 2) validate the calibrated model using additional toxicity data for pulsed exposure, and 3) derive profile-specific LP50 values (factor by which an exposure profile from a fate model must be multiplied to predict 50% mortality at the end of the profile).

A range of software implementations is available. The following implementations are included in a ring test that has been performed as part of the Cefic-LRI project (Jager and Ashauer 2018b), although more are likely to exist:

- Matlab implementation as a freely-downloadable package for BYOM: http://www.debtox.info/byom.html. This version is very flexible, but has limited user-friendliness (it requires a working knowledge of Matlab) and only a limited manual. This implementation is maintained by Tjalling Jager at DEBtox Research, The Netherlands.
- An R-package (freely-downloadable), the MORSE package, has been developed and is maintained by the University of Lyon, France: https://cran.r-project.org/web/packages/morse/index.html. A user-friendly web-based interface is available that includes the reduced GUTS models (http://pbil.univ-lyon1.fr/software/mosaic/survival/).
- A standalone version developed in Delphi; data are entered in the form of an Excel file. Developed and maintained by Thomas Preuss (Bayer). An earlier version can be downloaded as part of the supporting information of (Ashauer et al. 2016) at https://www.ecotoxmodels.org/guts/.
- EasyGUTS user interface for the GUTS R-package (Albert and Vogel, see below), developed by Dirk Nickisch (Rifcon, Germany). A test version for this software is currently available from Rifcon on request (see https://rifcon.de/downloads-2/).
- A Python toolbox for GUTS is available at GitHub (freely-downloadable): https://github.com/nepstad/epytox. It is developed and maintained by Raymond Nepstad (SINTEF, Norway). No manual, but an example notebook is included.
- GATEAUX, is a standalone Windows software built in C++ for Syngenta, based on the GUTS R-package (Albert and Vogel). This software was still in a beta version when used for the ring test, and will not be developed further.
- A Mathematica version has been developed by Andreas Focks (Alterra, The Netherlands). It can already be downloaded as part of the supporting information of (Ashauer et al. 2016) at https://www.ecotoxmodels.org/guts/. The author plans to make an updated version (with user manual) available for free download in the near future.
- A version in ModelMaker (a commercial general modelling platform) was developed by Roman Ashauer (Univ. York, UK, currently Syngenta, Switzerland). An implementation can already be

downloaded as part of the supporting information of Ashauer et al. (2016) at https://www.ecotox-models.org/guts/. This version will likely not be developed further.

- A version in OpenModel (a free modelling software) was developed by Roman Ashauer and Nina Cedergreen (Univ. Copenhagen, Denmark). However, at the moment, OpenModel does not include the multinomial likelihood that is needed for proper statistical treatment of the data (though the difference is usually small). The developer of OpenModel is interested, so this may be fixed in the future, and free distribution of this GUTS version is planned.
- The standalone implementation GUTS-3S was developed in Visual Basic by Judith Klein and Udo Hommen (Fraunhofer, Germany). This software is available for free download: https://www.ime.fraunhofer.de/en/Research_Divisions/business_fields_AE_BR/Businessareas_AE/Software_E/GUTS-3S.html, including a manual.
- The original DEBtox standalone Windows software was developed at the VU University (The Netherlands) over twenty years ago (Kooijman and Bedaux 1996a). This software is no longer offered or maintained, though it can still run on modern versions of Windows. It produces accurate parameter estimates, but only for the reduced hazard model of GUTS and is limited to fitting standard tests (constant exposure).

The first R-package that was developed for GUTS (freely-downloadable) can be found at https://cran.r-project.org/web/packages/GUTS/index.html. It has been developed by Carlo Albert (Eawag, Switzerland) and Sören Vogel, and is now being maintained by Rifcon (Germany). This package has limited user-friendliness as it requires a working knowledge of R, and only a limited reference manual is available (but see EasyGUTS). The developers did not participate in the ring test (although EasyGUTS applies this package as engine).

Recently, a Cefic-LRI funded project has delivered a frequentist-based and user-friendly standalone Windows software to perform the GUTS analyses (a Matlab version is available as well). This software follows the workflow as laid down in the EFSA opinion (openGUTS, see http://www.openguts.info).

Several software implementations apply a Bayesian framework rather than a 'frequentist' one. The difference between the two is briefly discussed in the GUTS e-book (Jager and Ashauer 2018b), as far as relevant to GUTS applications. In general, both approaches will deliver very similar results, and the choice between them is mainly a matter of taste. Bayesian statistics offers a more natural way to work with (and propagate) uncertainties, and probably yields more representative inference for small data sets (limited number of individuals). The price that needs to be paid is dealing with priors and the numerical difficulties of obtaining a representative sample from parameter space, particularly when one or more parameters cannot be identified from the data (run away to zero or infinity; several cases are illustrated in the "interpretation document" for openGUTS: http://openguts.info/download.html). In a regulatory context, it is essential to realise that the prior distributions of the parameters may exert an influence on the results; their appropriateness therefore must be evaluated as well. Frequentist applications generally apply hard minimum-maximum boundaries for the parameters, whose influence is easier to interpret.

2.2.2 Model Description

2.2.2.1 Problem Definition

Context in which the Model will be used

GUTS deals with the survival probability of individuals over time. It can thus be used to address questions that relate to the analysis of toxicity tests on mortality (or immobility), and the prediction of survival for untested exposure conditions. GUTS is not restricted to a particular tier; it can be used in Tier 1 to analyse data from standard acute tests, but also in higher tiers to analyse more complex non-standard data sets (e.g., with time-varying exposure), to predict survival for untested conditions, or as module in Higher Tier models (see list of possible applications in the general text above). The EFSA Sci Op. on TKTD Modelling (EFSA PPR 2018) focusses on application of GUTS in Tier 2C (predicting survival as a result of fate-model output).

Specification of the question(s) that should be answered with the model

Questions that relate to the survival probability of individuals, as function of time and exposure (the profile of concentration versus time, e.g., output from fate models).

Specification of necessary model outputs and protection goals

The model can be used to analyse (standard) toxicity data, estimate an LCx,t (for any effect percentage x, and any time point t), and predict survival probability as function of time for any exposure profile. The model is thus relevant for protection goals that deal with individual survival (e.g., for vertebrates), as well as cases where effects on survival are an important aspect of the population impacts (e.g., in combination with a population model).

Domain of applicability of the model

Analysing and predicting mortality (and immobility) in cohorts of individuals. The model is in principle applicable to all species. So far, it has been applied to animals only, but there is no reason why it should not be useful for other organisms as well (the toxicokinetics module is the main part that would need to be adapted to the species of interest). The applicability domain is mainly determined by the available data for calibration and testing (validation) of the model, and the extent of extrapolation that the model is used for (uncertainty increases with increasing distance between the situation for calibration and the situation for prediction).

Why is the model being used?

Because standard approaches (e.g., fitting dose-response curves) do not consider exposure time and exposure profile, and thus introduce considerable uncertainty in the risk assessment (Jager 2011), cannot accommodate non-standard toxicity data (e.g., when exposure is not constant), and cannot be used for meaningful predictions (e.g., for time-varying exposure profiles in the field).

What protection goal is being addressed?

Protection goals that relate to the survival probability of individuals. GUTS can also be used as individual-level module in models at higher levels of organisation (e.g., population models), and thereby aid in addressing other protection goals.

What outputs are required?

The model can be used to analyse (standard) toxicity data. Output are model parameters (including a time-independent threshold for effects) with confidence intervals. These model parameters can subsequently be used to estimate an LC*x*,*t* (for any effect percentage *x*, and any time point *t*), and predict survival probability as function of time for any exposure profile.

How was the species chosen?

GUTS can, in principle, be used for any species (though so far only data for animals has been used).

Which other species/groups are being covered by the chosen one(s)?

GUTS can, in principle, be used for any species (though so far only animals). If the model is calibrated to data for one species, the parameterisation reflects that species (and perhaps even only the life stage that was tested). There is currently insufficient information to identify patterns in parameter values across species, though some proof-of-concept was delivered for extrapolation of toxicity across life stages (see Gerritsen et al. 1998). This is similar to the limitations of a dose-response curve (or LC50 derived from it), with the remark that GUTS explicitly deals with effects over time and is thus better suited to identify the mechanistic basis underlying sensitivity differences.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

That depends on the specific application: it depends on the data that are being used and the question that is to be answered. For example, if GUTS is to be used to predict survival for an untested exposure profile, it would be useful to test the predictions of the calibrated model with a few additional toxicity tests (with another exposure profile). Validation with pulse-exposure toxicity tests is therefore required for aquatic risk assessment under the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018), for each case. The opinion proposes several goodness-of-fit measures, but no explicit pass-fail criteria are provided. However, even if the model is not properly validated according to the EFSA requirements, it can be defended that a GUTS prediction will still constitute a better-educated guess than predictions from an LC50 (which is also a model, albeit a very poor one).

2.2.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

GUTS is not built on data; the model is fully parameterised using toxicity data for a specific chemical-species combination. This is a very different situation than for population or fate models, which are generally not fitted (see Jager and Ashauer 2018a). In GUTS, data are only used for calibrating and validating the model in each application case. The data that are available for calibration and testing depends on the specific application that GUTS is used for, and will be specific for the species and compound under consideration. In general, calibration of GUTS requires data for survival in a group of test animals over time (several observation time points). The exposure concentration does not need to be kept constant (but must be known), and, in fact, if the model is to be used to predict mortality due to time-varying exposure, it makes sense to calibrate the model using data from tests with non-constant exposure (to minimise the distance for extrapolation). The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) proposes a range of requirements to data sets for calibration and validation for the purpose of aquatic ERA of pesticides.

Assessment of the quality of the data

That depends on the specific application and the specific species-compound combination that is considered; the model itself does not contain any case-independent data in any way. The currently-used quality controls on toxicity testing also apply here. GUTS is able to work with the results from most standard acute toxicity tests (as long as survival is scored at several points in time), as well as non-standard tests (e.g., tests where exposure is not kept constant). Of course, regulatory frameworks have their requirements for assessing the quality of toxicity data (e.g., criteria laid down in test guidelines). However, these criteria are not necessarily relevant for GUTS application (e.g., the demand for constant exposure in standard tests). The EFSA opinion provides a range of requirements that are more tailored towards TKTD models (e.g., for validation in Section 4.1.4.5), which should be applied for each model analysis.

2.2.2.3 Conceptual Model

Description of the model concepts including a diagram

See general text above, and flow diagram already presented in section 2.2.1.1. The two reduced cases of GUTS are most likely to be used in ERA (this is also recommended by EFSA). A first-order one-compartment model links the external concentration (as function of time) to the internal damage level. Damage over time is subsequently linked to the death mechanism. Two mechanisms are considered in the reduced models: death is stochastic at the level of the individual and all individuals are the same (SD), or death is deterministic at the level of the individual and individuals differ in their sensitivity (IT). Since it is difficult to select the most realistic representation of death, it is appropriate to use both models to cover the range of possible outcomes (and generally use the most conservative one for ERA).

Identify the main components and processes in the system

As with all TKTD models, GUTS is built up from a TK (toxicokinetic) and TD (toxicodynamic) module. The TK module is (by default) the one-compartment model with first-order kinetics. The TD model includes damage dynamics and the death mechanism itself. In the reduced models, TK and damage dynamics are combined in a single compartment with first-order kinetics.

How the effects of the chemicals are modelled

A one-compartment model is used to translate external concentrations to internal damage levels. The damage level affects the probability to die for each individual through two mechanisms: stochastic death (SD, damage above a threshold increases the individual's probability to die) and individual tolerance (IT, individuals differ in their threshold for effects).

How the components and processes are linked

A one-compartment model is used to translate external concentrations to internal damage levels. The damage level affects the probability to die for each individual through two mechanisms: stochastic death (SD, damage above a threshold increases the individual's probability to die) and individual tolerance (IT, individuals differ in their threshold for effects).

2.2.2.4 Formal Model

Identification of the model variables

For the reduced models, the state variables are the scaled damage level and the survival probability. Forcing variable is the exposure concentration (which may be time varying).

Identification of the model parameters

For the reduced SD model: dominant rate constant (k_d), median of threshold distribution (m_w), killing rate (b_w), and background hazard rate (h_b). For the reduced IT model, the killing rate is removed and the width of threshold distribution (β) enters as a new parameter. The reduced models thus have four parameters that need to be estimated from toxicity data (see example in section 2.2.1).

Description of the most important model equations or algorithms

Here, the equations will not be explained in detail; they are included to indicate the level of complexity of the model. For the reduced SD model:

$$\frac{dD_w}{dt} = k_d(C_w - D_w)$$

$$h_z = b_w \max(0, D_w - z_w) + h_b \text{ with } z_w = m_w$$

$$\frac{dS}{dt} = -h_z S$$

Where D_w is the scaled damage level, C_w is the external concentration, and S is the survival probability. For the reduced IT model:

$$\frac{dD_w}{dt} = k_d(C_w - D_w)$$

$$D_{wm} = \max_{0<\tau<t} D_w(\tau)$$

$$S = \int_{D_{wm}}^{\infty} f\left(z_{w;m_w,\beta}\right) d_z = 1 - F(D_{wm})$$

Where f is the distribution of thresholds, and F its cumulative distribution. D_{wm} is the maximum damage level over time until time point t. This extension is needed for situations where damage levels decrease in time (as usually happens when exposure varies over time), to avoid dead animals resurrecting.

2.2.2.5 Computer Model

Description of the model implementation

There is a range of model implementations available for GUTS, probably a few dozen. See also the list of model implementations in Section 2.2.1.2; these were the ones that were included in a recent ring test (Jager and Ashauer 2018b).

Checking the computer model for errors, bugs and inconsistencies in the code

Each implementation has its own development history. Several implementations have been extensively used and tested although these tests have seldom been formalised and documented at this point. The recent ring test compares a range of implementations in a more formal manner. The EFSA opinion

recommends that each new implementation is tested against the ring-test data to confirm that it is working properly. Limitation of this collection of test data is that it does not include a proper case of 'slow kinetics' (dominant rate constant k_d running to zero). Such cases are tough on the numerical methods, and tend to clarify the influence of any priors or minimum bounds to parameters. For testing openGUTS, the ring-test data were extended by series of other data sets (including 'slow kinetics', see http://openguts.info/download.html).

Demonstrate that the computer model performs as indicated by the conceptual and formal models

Each implementation has its own development history. Several implementations have been extensively used and tested although these tests have seldom been formalised and documented at this point. The recent ring test compares a range of implementations, and indicates their robustness (as well as some limitations for specific implementations).

2.2.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context **in which the model is run**

For GUTS, the scenario entails the exposure profile (i.e., the exposure concentration as function of time). This scenario should thus be defined for each calibration analysis (from the design of the toxicity test) and for each extrapolation (e.g., the exposure profile from a fate model). One thing to consider is that GUTS parameters may well depend on the environmental temperature; for example, we can expect rate constants to increase with an increase in temperature. Furthermore, several parameters (such as TK rate constants) may depend on body size or life stage of the individual. Uncertainty in model predictions will increase with increasing difference between the conditions used for the toxicity tests for calibration and the conditions envisaged for the predictions.

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

The values of all GUTS parameters are established by fitting the model to a toxicity data set, usually conducted under standardised laboratory conditions. It is good to realise that when GUTS is used for predicting mortality under time-varying exposure profiles, the pesticide concentrations will have been generated using the relevant exposure models. These exposure models will include considerations of environmental factors such as soil type, rainfall and agronomic practice.

2.2.2.7 Parameter Estimation

Description of the model parameter estimation

The values of all GUTS parameters are established by fitting the model to a toxicity data set (numbers of survivors over time in various treatments) for each case. There are no parameter values that can be considered 'part of the model'. GUTS can be calibrated using the data from many standard test protocols, as long as observations on survival are reported at several points in time. However, it can also be calibrated on non-standard data sets, for example, data sets where the exposure concentration was not kept constant over time. It should be noted that standard test protocols have not been optimised for the purpose of calibrating mechanistic models; they were optimised for fitting a dose-response curve on the results at the end of the test to derive an LC50. Other test designs will likely be far more efficient for the modelling purpose. Optimal test design will depend on the specific properties of the test chemical, as well as on the question to which the model is applied (Jager 2014). However, in general, it is a good idea to include more observations over time and flexibility in test duration (extending a test when there is little mortality, or mortality is only slowly increasing over time). Furthermore, it may

well be that calibration on time-varying exposure is more informative than on constant exposure. Simulation studies will be needed to study optimal test design for TKTD models, for different purposes (see e.g., Albert et al. 2012).

Parameters estimated from the literature — what are the sources and why are these appropriate?

No parameters are estimated from the literature; all parameters will be fitted on the survival data over time (though these data themselves may have been extracted from publications). Exception can be the background hazard rate, which may be set to zero, or to a general value for the test species under the test conditions.

Parameters obtained from calibration — how and why this was done?

The values of all GUTS parameters are established by fitting the model to a toxicity data set. This is done by maximising a likelihood function based on the multinomial distribution. Different software implementations use different numerical schemes for optimisation (e.g., Bayesian inferences or Nelder-Mead simplex optimisation).

2.2.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

Sensitivity analysis is pointless for models that are completely parameterised by fitting them to data; this holds for GUTS as well as for dose-response curves (Jager and Ashauer 2018a). The relevant information on sensitivity and identifiability is contained in the (joint) confidence interval of the parameters as a result of the calibration. If required, a classical sensitivity analysis *can* be performed, but it would have to be done after the model is calibrated to a specific data set, in each specific case (Jager and Ashauer 2018b). The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) includes a sensitivity analysis using hypothetical parameter values as a reference and a rather arbitrary time-varying exposure scenario. The purpose of this analysis is not so clear, although it shows that all parameters are relevant, and it could be used as one of the checks for the correctness of a model implementation.

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

In classical uncertainty analysis, all parameters receive an (independent) distribution, which is propagated to the model output. This is done once (or several times during development) for the model by the developer and serves as a 'mark of model quality' (EFSA PPR 2014b). This is also a rather pointless exercise for models that are completely parameterised by fitting them to data (Jager and Ashauer 2018a): we would need to give all parameters a distribution between zero and infinity as they are completely dependent on the data set and thus case-specific. The result of such an analysis would be meaningless. Instead, the uncertainty in the parameters should be taken from the model fit, which implies that an uncertainty analysis has to be done in each case, for each fit. In general, such a procedure is not called an uncertainty analysis but referred to as error propagation.

Uncertainty in the model parameters resulting from a fit can be propagated to uncertainty in the model predictions, for each case, using various methods (Bayesian or frequentist based). This is a very important analysis as it shows the impact of the uncertainties in the parameter identification (even though uncertainty in the exposure profile for the predictions will also play a role, which is generally ignored). Part of the structural uncertainty can be addressed by using both the SD and the IT model to

the same data (the 'truth' is likely in between these two extreme views). Other factors that cause uncertainty in model predictions are discussed qualitatively at the end of this evaluation.

2.2.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

This item cannot be addressed for the model in general; the model output can only be compared to independent data after the model has been calibrated using data for a specific species-chemical combination. Such comparisons thus can only be performed as part of a specific application (dossier), where they can increase confidence in the model and its parameterisation for the specific case at hand. For models that are necessarily fitted to data, comparison to independent data only makes sense when the model is used in extrapolation; e.g., when GUTS is calibrated to data for constant exposure and used to predict survival due to pulsed exposure. A few published examples of such comparisons are available (Ashauer et al. 2007b, Nyman et al. 2012, Ashauer et al. 2016, Focks et al. 2018). The workflow laid down in the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) explicitly specifies a validation step with pulsed-exposure testing for each species-chemical combination.

Demonstration that the model output provides an adequate match to data patterns

Survival models have been used for many decades and generally provide a good fit to survival data over time. For SD models, a list of publications is maintained at http://www.debtox.info/papers_survival.html (at this moment, over 100 papers). There are currently no alternative models that even come close to the same explanatory (or descriptive) power with the same low number of free parameters.

2.2.2.10 Model Use

Explanation of how the model conforms to the requirements set in the problem definition

This item cannot be addressed in general. GUTS should be seen as dose-response model, but then more robust and more mechanistic than descriptive methods (such as fitting a log-logistic dose-response curve at one time point, or hypothesis testing to derive a NOEC). Because of this mechanistic nature, a calibrated GUTS model can be used to make predictions for untested (time-varying) exposure situations.

Description how the model works (user manual).

A full model description is available in the open literature (Jager et al. 2011) and more detailed in an e-book (Jager and Ashauer 2018b) (see also http://www.debtox.info/book_guts.html). Many of the implementations have some form of manual, or one is being planned.

Description of the pesticide parameters values used in the model

The values of all GUTS parameters are established by fitting the model to a toxicity data set for the specific pesticide in each case. The model itself contains no data whatsoever.

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed in general.

2.2.3 Model Evaluation

Note that this section has been filled according to Appendix B, summary checklist for model evaluation by the risk assessor, proposed in the EFSA Sci. Op. on GMP (EFSA PPR 2014b). Since then, the recommended checklist has been adapted to GUTS models (EFSA PPR 2018).

2.2.3.1 Problem Definition

The regulatory context in which the model is run

GUTS deals with the survival probability of individuals over time. It can thus be used to address questions that relate to the analysis of toxicity tests on mortality (or immobility), and the prediction of survival for untested exposure conditions. GUTS can be used for various purposes, as explained in the general text above. The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on application of GUTS in Tier 2C.

The question that has to be answered with the model

Questions that relate to the survival probability of individuals of a specific species, as function of time and exposure profile. GUTS is limited to effects on the endpoint mortality (and possibly other quantal endpoints such as immobility, as long as effects can be treated as non-reversible). Furthermore, GUTS is a model for individuals. However, the model output is the probability for an individual to die, as function of time, and depending on the exposure pattern. Therefore, it is fitted to, and can predict, survival of cohorts over time. Furthermore, it may be (and has been) implemented into higher-level (e.g., population) models.

The available knowledge and data relevant to the risk assessment question

This cannot be answered for the model in general. GUTS unifies scientific principles that have a long tradition in ecotoxicology (one-compartment TK, critical body residues, hazard modelling, etc.).

The outputs required to answer these questions including performance criteria for the regulatory model

This cannot be answered for the model in general. GUTS can produce various types of useful outputs such as a threshold for effects on mortality, LC$x,t$ for any effect percentage x and exposure duration t, or the expected survival pattern for any (untested) exposure profile. The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018), focusses on the latter application, and proposes a number of criteria for the data and model performance in calibration/validation. The calibrated model needs to be validated for each species-chemical combination, against a series of pulse-exposure treatments, before it can be used for extrapolation to exposure profiles from fate models.

The species to be modelled

GUTS can be used for many, in principle all, species and all chemicals (as well as 'non-chemical' stresses such as ageing and microplastics). However, for some chemicals and species, model adaptations may be needed. For example, species that build up substantial lipid storage may require a two-compartment TK model for hydrophobic chemicals (Jager et al. 2017a). GUTS may in the future support extrapolations between species, but at this moment, such extrapolations have an unknown degree of uncertainty as only few comparative studies have been performed.

Requirements for the environmental scenarios to be used in the risk assessment

The model does not rely on an 'environmental scenario' apart from the exposure profile (how the exposure concentration varies over time, e.g., the output from fate models). However, environmental conditions, such as temperature and food availability, may influence parameter values of the model. Little is known about the effects of environmental conditions on GUTS model parameters at the moment. When GUTS is used to predict effect as a result of fate models (e.g., FOCUS output), the fate models will require environmental scenarios that consider factors such as soil type, rainfall and agronomic practice.

2.2.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

This item cannot be answered for the model in general. The model itself contains no data; all model parameters are calibrated on toxicity data for a specific chemical-species combination, and are thus case specific. Standard acute toxicity data can be used (when survival is reported at several time points), although the standard protocols have not been optimised for parameterising GUTS (Jager 2014).

Has the quality of the data used been considered and documented?

This item cannot be answered for the model in general. The only data that are used by the model are toxicity data for a specific chemical-species combination, and thus case specific.

Have all available data been used? If not, is there a justification why this information has not been used?

This item cannot be answered for the model in general. The only data that are used by the model are toxicity data for a specific chemical-species combination, and thus case specific.

2.2.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

This item cannot be answered for the model in general.

Are the modelling endpoints relevant to the specific protection goal?

This item cannot be answered for the model in general. The endpoint of GUTS is mortality, or more specifically: the survival probability of a cohort over time, as a function of an exposure pattern in the environment. The model is thus relevant for protection goals that involve mortality (e.g., vertebrates). Mortality is clearly an important aspect of population dynamics, although not necessarily the most sensitive one (in general, sub-lethal endpoints will be affected at lower exposure concentration than mortality). GUTS has been included in several population models to represent the survival of individuals as a function of time and exposure concentration.

Is the modelling approach justified?

GUTS unifies all of the mechanistic survival models that have been used in (eco)toxicology into a single framework. Specific models that have been used (and are being used) can be seen as special cases of GUTS. This implies that there is no alternative for GUTS (when it comes to TKTD models for survival with few parameters), and that it is a justified approach for modelling survival. The EFSA Sci. Op. on

TKTD Modeling (EFSA PPR 2018) considers GUTS "ready to be used in risk assessment." GUTS is meant to replace descriptive methods for dose-response analysis.

Is the conceptual model logical?

The model concept is simple and logical, as explained in section 2.2.1. In most situations, one or more special cases from the GUTS framework will be fitted to data; specifically the reduced models that only require four parameters (see example case study in section 2.2.1). Given the type of data sets that are available for calibration (which generally does not include information on body residues or sub-organismal effects), this is the maximum level of model complexity that can be accommodated.

Are the processes included in the model relevant to the addressed issue?

Yes; see section 2.2.1 at the start of this chapter.

Are the links between different processes to the variables logical?

Yes; see section 2.2.1 at the start of this chapter.

Are the temporal and spatial scales relevant in regard to the problem definition?

GUTS has no spatial scale. In terms of temporal scale, there are no restrictions. However, for most applications it is assumed that the model parameters remain constant over the time period modelled. This is generally valid for acute toxicity tests, but becomes more questionable for longer test durations and for extrapolation to longer exposure times (which is the application foreseen in the EFSA opinion). When organisms grow and develop (or starve), their model parameters may well change. For example, increase in body size will affect toxicokinetics by growth dilution and by changing the surface:volume ratio. In general, uncertainty increases with increasing distance between the extrapolation scenario and the conditions of the toxicity test(s) used for calibration. Uncertainties and limitations are discussed in more detail at the end of this chapter.

2.2.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

GUTS was first presented in a 2011 paper (Jager et al. 2011), including a discussion of the concepts and clarification of the model equations and parameters involved. No new model concepts were introduced at that point; all of the components of GUTS had already been used in the scientific community for many decades. GUTS simply unifies these previous models into a consistent over-arching framework. Recently, an e-book on GUTS has been released (Jager and Ashauer 2018b) as part of a Cefic-LRI funded project (http://debtox.nl/projects/project_guts.html). This e-book includes an extensive description of GUTS (both conceptual and mathematical, including an explicit explanation of the underlying assumptions), case studies, and detailed guidance of how to apply the model in practice.

Are the most important mathematical equations described?

Yes, see previous point.

Is there a description of the variables and parameters including their meaning and unit?

Yes, see previous point.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

GUTS is specifically intended to be applied in the absence of detailed information on the toxicity mechanism of the compound and the biology of the species. The simplest cases of GUTS can be calibrated using results from standard acute toxicity tests (number of survivors over time). The more complex cases of GUTS (e.g., those including a separate toxicokinetic and damage module) additionally require information about body residues over time. GUTS represents the maximum level of complexity that is still useful for such types of data sets.

Are references supporting the equations been provided?

Yes, see above.

2.2.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

There are many implementations of GUTS. In a ring test, recently conducted in the framework of a Cefic-LRI project (Jager and Ashauer 2018b), 11 (largely) independent software implementations have been compared (see list in Section 2.2.1.2). These implementations differ in terms of their user-friendliness, description, code availability, and verification status. The detailed ring testing provides more clarity on the robustness of these implementations. Most of these implementations are freely available, or are planned to be made freely available in the near future, and some include user manuals. As an example, the BYOM platform for Matlab includes a dedicated package for GUTS applications: http://www.debtox.info/byom.html. This platform and the package are freely downloadable, but Matlab is a commercial program. BYOM includes a basic manual, but user friendliness is limited (there is no graphical user interface, the user has to work in the Matlab environment).

Is the computer code well readable and is it available?

This differs between the various implementations. For several implementations code is available and readable.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

This differs between the various implementations. For the Matlab implementation, for example, a range of checks has been performed to verify the correctness of the package, such as comparison to the DEBtox Windows software (Kooijman and Bedaux 1996a), comparison between the analytical solutions and the ODE solvers, and consideration of trivial cases. However, these verification attempts have not been documented. The ring test that was recently conducted sheds more light on the robustness of the various implementations; these results are documented in the e-book (Jager and Ashauer 2018b). The EFSA opinion recommends that each new implementation is tested against the ring-test data to confirm that it is working properly.

2.2.3.6 The Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

This item cannot be answered for the model in general. For GUTS, the 'scenario' entails the pattern of chemical concentration in the environment that the individuals will experience over time. The scenario will thus generally follow directly from the design of the toxicity test that is to be analysed, or from the environmental setting that needs to be simulated (e.g., using the output from a fate model).

GUTS parameters may depend on environmental conditions such as temperature and food availability, as well as on properties of the organism (body size, developmental status). At this moment, the effect of these factors on the model parameters has not been thoroughly studied. Therefore, extrapolations will become more uncertain with increasing distance between the extrapolation scenario and the situation of the experimental data used for calibration.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

This item cannot be answered for the model in general. For GUTS, the 'scenario' entails the pattern of chemical concentration in the environment over time, which is not part of the model itself.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

This item cannot be answered for the model in general. For GUTS, the 'scenario' entails the pattern of chemical concentration in the environment over time, which is not part of the model itself. GUTS has no spatial context (it is a model for individuals) and hence there is no 'area' to consider. For the animals, the exposure pathways included are the ones that are available in the toxicity test. In general, animals are not fed in standard acute toxicity tests, and hence uptake with food will not be considered in the model parameterisation nor in predictions.

Is the level of conservatism placed into the scenarios appropriate?

The scenario is not part of the model, but of a specific application. The model itself is not conservative or non-conservative, it is the scenario (the exposure profile over time) that will determine the level of conservatism (and should lead to selection of an appropriate assessment factor to cover the remaining uncertainties).

2.2.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

This item cannot be answered for the model in general. All model parameters obtain their value by fitting the model to toxicity data for a specific species-chemical combination. Therefore, parameter estimation is completely case specific.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

This item cannot be answered for the model in general. All model parameters obtain their value by fitting to the model to toxicity data for a specific species-chemical combination. GUTS can be calibrated on data resulting from standard test protocols for acute toxicity (when survival is scored at several time points). However, such toxicity tests have not been optimised for this purpose; non-standard test designs will likely be much more efficient (Jager 2014).

Were the estimated parameter values realistic?

This item cannot be answered for the model in general. All model parameters obtain their value by fitting the model to toxicity data for a specific species-chemical combination. A large number of GUTS analyses have been performed, with a range of species and toxicants, which could be used to generate 'expected ranges' for parameter values (see e.g., Jager and Kooijman 2009, Ashauer et al. 2015).

Are the data sources sufficiently documented?

This item cannot be answered for the model in general. All model parameters obtain their value by fitting the model to toxicity data for a specific species-chemical combination.

2.2.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

GUTS will be always fitted to toxicity data, which is specific for a chemical and a species. All model parameters obtain their value by fitting the model to a data set, and classical sensitivity and uncertainty analysis thus make little sense (just as those analyses do not make sense for the log-logistic dose-response curve). All of the relevant information on uncertainty and sensitivity is represented in the parameter estimates and their (joint) confidence intervals.

Sensitivity and uncertainty analysis *can* be performed for TKTD models like GUTS, for any analysis, after the model has been calibrated to a specific data set. However, there are very few examples of such analyses as their usefulness is very limited (Jager and Ashauer 2018a, Jager and Ashauer 2018b). The EFSA opinion includes a general sensitivity analysis using rather arbitrary reference values for the parameters and a rather arbitrary exposure situation. Instead of classical uncertainty/sensitivity analysis, it is more useful to focus on a proper statistical treatment for optimisation, and for the subsequent construction of confidence intervals, and the propagation of uncertainties.

If classical uncertainty/sensitivity analysis is required, one would first have to define the relevant model output. For GUTS, that may be the survival probability. However, the survival probability is a function of time and exposure pattern, and therefore, the sensitivity of parameters will be a function of time and exposure pattern as well. In general, all parameters will be sensitive, and contribute to uncertainty, at some time point under some exposure scenarios. An example of a rather classical sensitivity analysis for GUTS, in the case of an extrapolation to a FOCUS scenario, is provided in Ashauer et al. (2013). However, it is not so clear how this information, varying one parameter at a time across its confidence interval, would benefit a risk assessment; all parameters are fitted together, often correlated, and it is not possible to improve one of the parameters independently from the others (although it may be possible to use model simulations to optimise experimental design to have more resolution on a specific parameter).

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

Sensitivity analysis is hardly useful for models that are completely parameterised by fitting them to data.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

Sensitivity analysis is hardly useful for models that are completely parameterised by fitting them to data. We are able to select the most sensitive parameters (although this will depend on exposure time

and concentration profile), but there is little opportunity to refine one of them specifically (additional toxicity testing will be useful to refine parameters, but this will affect all of them).

Has the uncertainty analysis been adequately documented?

Classical uncertainty analysis is hardly useful for models that are completely parameterised by fitting them to data (some examples are provided in Jager and Ashauer 2018b). Uncertainties in the model parameters (resulting from the fitting) can be propagated to uncertainties in the model predictions (see below).

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

This item cannot be answered.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

Uncertainty analysis is hardly useful for models that are completely parameterised by fitting them to data. We are able to select parameters that contribute most to the uncertainty (although this will depend on exposure time and concentration profile), but there is little opportunity to refine one of them specifically.

Uncertainty is propagated to the model results?

Uncertainty analysis is best covered by propagating the (joint) uncertainty in the parameters estimates to the model predictions, which can be performed by several of the implementations currently in use (see example in Ashauer et al. 2016). Such an analysis will only include the uncertainty in the model parameters as estimated from the data. Part of the conceptual uncertainty can be visualised by fitting both SD and IT cases of GUTS to the same data, and making predictions from these calibrated models. However, there will be additional uncertainties (see specific text at the end of this evaluation), which will become more prominent with increasing extrapolation distance. These uncertainties will need to be addressed with additional experimental work and/or an appropriate assessment factor.

It is good to stress that part of the uncertainty in the model predictions will derive from uncertainties in the exposure profile (i.e., the output from the fate models). For a meaningful and consistent propagation of uncertainties, this would have to be done throughout the risk assessment (making the risk assessment probabilistic), and not exclusively for the effect models.

Have confidence intervals been estimated and has this information been used in further model use?

When fitting the model to the data, confidence intervals on the model parameters are derived by most implementations of GUTS (either based on a frequentist or Bayesian framework). The joint confidence intervals on the parameters can be propagated to obtain intervals on model predictions, which can be an LCx,t or the expected survival probability due to an untested exposure profile.

2.2.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. For each case, specific performance criteria may be defined. The ESFA Sci. Op. on TKTD Modelling (EFSA PPR 2018) proposes several criteria for application in ERA of PPPs, a number of requirements

for the validation data and a number of model-performance criteria (qualitative and quantitative, though no strict pass/fail cut offs are provided).

Are the model outputs that are compared relevant in view of the problem definition?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. Model validation should be closely linked to the intended purpose of the model, and as GUTS can be used for different purposes (see Section 2.2.1.2), this item cannot be addressed in general. If GUTS is to be used to fit experimental data and derive a no-effect concentration or LC*x*,*t*, validation is impossible. The only criterion for 'validity' can then be goodness-of-fit, and whether the model is generally able to provide a good fit to survival data (the same is true for dose-response curves and TK models). If GUTS is to be used to extrapolate from one exposure scenario to another (e.g., from constant to a pulsed exposure scenario), validation *is* possible. A few of such studies have been performed (Ashauer et al. 2007b, Nyman et al. 2012, Ashauer et al. 2016, Focks et al. 2018), but more structural validation work would be beneficial. It should be noted that such a validation study, in principle, only says something about the validity of the model parameterised for *that* particular species and chemical (and *that* particular extrapolation). Therefore, the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) requires a validation for each case. However, a range of such validations goes a long way towards embodying trust in the general model structure, and can provide information on the accuracy/precision of extrapolations, which in turn may be used to set reasonable assessment factors. Ultimately, this can lead to modification of the validation requirements for each application.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. The validation studies that have been performed are mostly published, with details on the data used (see above).

Is the dataset relevant in view of the problem definition?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific.

Is the fit of model output to the data good enough?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. Furthermore, it is unclear what 'good enough' means in this context. Regarding the validation studies that have been published, and looking at extrapolation from constant to pulsed exposure, the predictions are reasonable (and worst case) in two studies (Ashauer et al. 2007b, Nyman et al. 2012) but not so convincing in another (Ashauer et al. 2016). However, it should be noted that these studies were performed with field collected animals (*Gammarus pulex*), which may have added noise.

Has the performance of the model been reported in an objective and reproducible way?

In the published validation studies, plots were made to judge the performance of the predictions visually. In one study (Ashauer et al. 2016), prediction intervals were included in these plots, which helps to judge the deviations in view of the uncertainties in the model parameters and the effects of a small

population in the experimental test (since death is treated as a chance process, mortality in a small population may strongly deviate from the expectations simply due to the randomness of the effect). EFSA proposes a number of model-performance criteria for validation in GUTS applications.

2.2.3.10 Model Use

Is a user manual available?

Some implementations of GUTS have a user manual, although more work would be needed if the implementations are to be used by non-specialists. However, there is a lot of development going on in this area at the moment, with various groups developing software for GUTS (several of them focusing specifically on application in the ERA context, such as openGUTS, which was released in December 2019).

Have all aspects of the modelling cycle been documented?

For models like GUTS, there is no modelling cycle. Such a cycle is an (overly) simplified representation, that most closely matches the situation where a model is built from scratch by a single research group (or a single person), for a specific purpose (Jager and Ashauer 2018a). GUTS is a unification of many different models, each with its own history. GUTS has several special cases, several dozen different software implementations, and probably more than a thousand applications for different species-chemical combinations. Dozens of research groups have worked (largely independently) on survival models that are now seen as special cases of GUTS, and have applied these models for many different purposes. Furthermore, *all* GUTS model parameters receive their value in a case-specific calibration. Therefore, parameter-dependent steps in the modelling cycle such as sensitivity/uncertainty analysis and validation can only be performed for specific cases, and cannot be considered as part of a general modelling cycle.

Has a summary sheet been provided by the modeller?

This item cannot be answered for the model in general. The previous table is basically the summary sheet for the model in general.

2.2.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

We cannot say whether the modellers who will produce a GUTS analysis for a particular dossier are available for dialogue. Many scientists are working on the general development and application of GUTS, and many of them would likely be open to dialogue with risk assessors.

Is a version control system implemented?

GUTS itself does not have a version control. Basically, there is only one version. Although the presentation of the model has slightly changed in the recent e-book (Jager and Ashauer 2018b), the model remains identical mathematically. It should be noted that GUTS is not a proprietary model, the name is not protected, and there is no central control over its development. Many researchers around the world are working on survival modelling, and anyone can make a modified version and call it GUTS. However, it is conceivable that the recently-published e-book will become the standard work for GUTS. Several of the software implementations have a version control system, as these are more liable to be updated once in a while.

2.2.3.12 **Overall Judgement**

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

In general, GUTS was judged to be suitable for regulatory purposes by EFSA (EFSA PPR 2018). However, suitability of a model analysis will need to be evaluated in the context of a specific risk assessment (for a specific chemical-species combination), a specific use of the model, and after suitability criteria have been laid down. Furthermore, in general, suitability for ERA needs to be considered in view of the available alternatives for the modelling.

GUTS can be used for analysis of toxicity data, as a robust replacement of descriptive dose-response curves (or 'model-free' Spearman-Karber). From a scientific perspective, GUTS is far more suitable for this task than the available descriptive alternatives as it explicitly deals with the time dependence of toxicity. Furthermore, a calibrated GUTS model can be used to predict survival as a result of a different exposure scenario (e.g., the output from a fate model). These extrapolations come with uncertainties; uncertainty will increase with increasing distance between the exposure situation in the calibration data and the extrapolation scenario. For example, extrapolation from a 4-day standard test with constant exposure to a 1-year time-varying exposure profile is more uncertain than extrapolation to a 10-day constant exposure scenario.

An appropriate safety margin would be needed to address these uncertainties. Additionally, applicants could shed more light on these uncertainties by performing additional toxicity tests, aimed at clarifying specific aspects (e.g., toxicity tests with pulsed exposure, or at different temperatures). Such additional tests could be used to test the predictive abilities of the calibrated model (i.e., validation), to provide a more robust calibration (fitting on all available data sets together), or to extend the model (e.g., use a more complex case from the GUTS framework). Testing the calibrated GUTS model against pulse-exposure data is currently a requirement formulated in the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018).

Despite these uncertainties, GUTS is far more suitable for this extrapolation task than the LC50 and the time-weighted or peak exposure concentration, as it is based on a well-established mechanistic representation of TK and TD. Note that the use of an LC50 is also a model, which shares the uncertainties listed above, and adds a few more (especially those related to time/timing of exposure). Furthermore, the GUTS model has a broad support in the scientific community, and unifies the work on survival modelling over at least half a century. Therefore, the model is, in general, a good candidate to consider for regulatory purposes, and as such has been judged "ready for use" by EFSA (EFSA PPR 2018). However, more structured scientific work would be helpful, to demonstrate the accuracy and precision of GUTS in extrapolations, which would aid the selection of appropriate assessment factors (and ultimately refine the requirements for validation tests).

GUTS is a useful addition to many population models (especially IBMs) to deal with survival effects of individuals as a function of their exposure history. For this purpose, the model is far more useful than a static dose-response curve. However, the population context requires some more considerations, as individuals grow and develop in such models (which might influence model parameters, and includes death by old age or predation), and as chemicals may be transferred from mother to offspring. Furthermore, additional stresses in the population context (e.g., intra-specific interactions or food limitation) may affect the GUTS model parameters; such interactions cannot yet be predicted.

2.2.4 Qualitative Assessment of Uncertainties

GUTS models are not by themselves conservative or non-conservative. The level of conservatism depends on how the model is used, which data are used to calibrate it, which model outputs are used, and what assessment factor is selected. We can, however, identify certain areas that will affect the level of conservatism, which are presented below. An appropriate assessment factor will be needed to ensure that the overall level of conservatism is acceptable. It is good to note that the same uncertainties apply to the methods currently used to deal with mortality in ERA (dose-response analysis, and using LC50 in conjunction with a time-weighted average or peak concentration). These methods are also models, and should be evaluated along the same lines as the GUTS model (Jager and Ashauer 2018a). However, the current LC50-based analyses have a range of additional uncertainties that will lead to far more severe under- and overestimation of risk (especially due to ignoring the time aspect of toxicity). GUTS extracts the maximum amount of information from acute toxicity data, but this type of data is inherently limited in the amount of information that it contains. Mechanisms and processes that are not observed in the experimental test would be difficult or even impossible to predict.

2.2.4.1 Potential for Underestimation of Real Risk

- Short toxicity tests may not reveal additional effect mechanisms of the chemical if these mechanisms only show up after prolonged exposure to low concentrations (some studies indicate the existence of such additional mechanisms). The model cannot be used to predict such long-term effects from short-term toxicity tests. Extrapolation to longer test durations thus rests on the assumption that the mechanisms observed in the short-term test are also the only ones relevant for the extrapolation scenario.
- Some chemicals may specifically affect mortality during a particular stage of the life cycle (e.g., during embryonic development or metamorphosis). If such a stage was not represented in the toxicity test, its specific sensitivity cannot generally be identified, and the model will not predict it.
- Acute toxicity tests are performed in absence of food, but otherwise optimal conditions in terms of temperature, absence of predators or diseases, absence of other toxicants, etc. As GUTS focusses on direct mortality due to chemical stress only, it will not be able to predict potential synergism with other stresses (e.g., if a chemical makes a species more prone to disease or predation).
- GUTS only applies to mortality (and immobility). The chemical may affect growth and reproduction at (much) lower concentrations, which may have strong effects on population dynamics. For effects on growth and reproduction, other models (such as DEBtox) would be needed, and other types of toxicity tests (i.e., partial life-cycle tests). It is good to note that DEBtox models often include a survival model (usually the GUTS SD model) to perform an integrated analysis of lethality and sub-lethal endpoints.

2.2.4.2 Potential for Overestimation of Real Risk

- Acute toxicity tests are generally performed in the complete absence of food. Starvation may exacerbate the effects of toxicants. If such starvation events are not relevant for the field situation, overestimation is possible. However, the duration of acute toxicity tests is generally short enough to minimise negative effects due to starvation.
- When organisms are not fed in a test, they will not grow. Growth tends to dilute internal concentrations of chemicals, thereby delaying their action, and enhancing the effective 'depuration' when exposure ceases. Feeding may also affect the toxicokinetics of the chemical (especially the speed at which steady-state is achieved), so the consequences of feeding may be quite complex.
- Acute toxicity tests are usually performed with juveniles. If juveniles have a differential sensitivity, they are usually found to be more sensitive than adults. Furthermore, the small body size implies a

large surface:volume ratio, which implies that chemicals are taken up faster than in larger conspecifics, revealing the toxic effects earlier after the start of exposure. Also, juveniles do not reproduce, and reproduction could serve as an effective elimination route in adults.

- GUTS only applies to mortality of individuals. When extrapolating to longer exposure duration, the prediction is for a single cohort, exposed for the entire time period. There is thus no consideration of recruitment by reproduction or immigration from other areas, which would require a population context.
- GUTS does not include explicit consideration of bioavailability. In many cases, bioavailability of chemicals will be lower in a field situation than in a toxicity test. Bioavailability has to be dealt with separately.

2.2.4.3 **Potential for Uncertainty in Either Direction**

- Extrapolations with GUTS will rest on the assumption that the parameters remain constant at the value established in the calibration. However, several parameters will be affected by temperature (especially rate constants), body size/growth (especially TK rate constants), and presence of other stressors (possibly the TD parameters such as the threshold).
- Linked to the previous point, constancy of parameters also implies that effects are assumed to be reversible when exposure stops, with a time constant that is the same as in the accumulation phase (the time to reach steady state with the exposure concentration is the same during accumulation and depuration). For some chemicals, there could be hysteresis in the dynamics of the effects. For example, some forms of damage might not be completely repaired in a clean environment, or organisms may have inducible defences that render them less sensitive to subsequent exposure events.
- Calibrating GUTS for species A and compound B does not imply that the same parameters also hold for other species and other chemicals. The model cannot be used to predict the results for other species than the test species, or other chemicals than the test chemical. Therefore, at this moment, the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) recommends calibrating and validating the GUTS models for each species-chemical combination. Several studies have indicated that the model offers great potential for inter-species and inter-chemical extrapolations, but these methods are not yet at a stage where they can be applied in ERA of pesticides.
- Linked to the previous point: laboratory strains of species may differ in sensitivity from individuals in the field (both in average sensitivity and in variation in sensitivity).

2.3 DEBtox

Evaluation by Tjalling Jager

2.3.1 General Information

2.3.1.1 Background and Concept

The first thing to clarify is that many models have been referred to as 'DEBtox' over the years. The name was first used for a standalone Windows software, with associated booklet (Kooijman and Bedaux 1996a), that contained simplified Dynamic Energy Budget (DEB) models to analyse standard toxicity tests for fish growth and Daphnia reproduction, as well as a hazard analysis for survival data (which is currently considered as special case of GUTS), and a simple population growth model for algae toxicity. The underlying models were published in a series of papers in the open literature in 1994/1996. More recently, the term DEBtox has been used more generally: for the application of DEB-models to (eco)toxicology. Some authors also refer to hazard-based survival models as DEBtox models, but this use of the term should be discouraged as survival models do not contain energy-budget considerations. Here, the focus lies on the last incarnation (at the time of writing) of the simplified DEB model 'DEBtox', namely the one presented by Jager and Zimmer (2012)[4], in which a few errors in previous derivations were corrected, and a coherent statistical framework was presented. Later in this section, related models will be presented together with the main differences with the version by Jager and Zimmer.

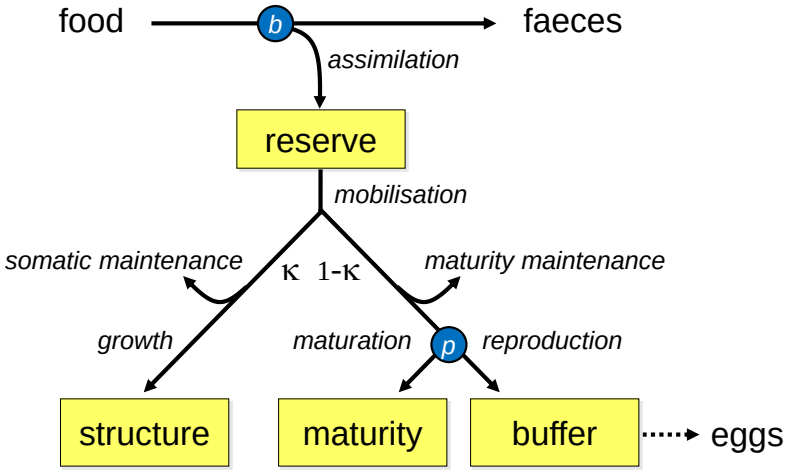
DEBtox has its basis in DEB theory (Nisbet et al. 2000, Kooijman 2001, Jusup et al. 2017), a coherent framework for bioenergetics of all life. The concepts behind DEB theory and its application to ecotoxicology have been presented in an extensive e-book (Jager 2019)[5]. This book provides a readable, math-free, introduction to DEB theory for animals, the extensions needed to deal with toxicant effects, case studies, etc. (the mathematical details are provided in a separate document). Here, a summary will be provided of the most important aspects of DEBtox.

DEBtox is based on the standard DEB model for animals, as formally presented by Sousa and co-workers (Sousa et al. 2008, Sousa et al. 2010). The structure of this model is shown in Fig. 4. The purpose of DEB theory is to provide a set of simple rules for how organisms take up resources from the environment, and how they use them to grow, develop and propagate. These rules follow the conservation laws for mass and energy, and DEB thus constitutes a true energy budget. Such a simple explanation for an individual's life history is useful for many different purposes, such as basic ecology, aquaculture and fisheries, climate change, and ecotoxicology (which is the focus here). A list of more than 550 papers on DEB has been compiled, which provides a good demonstration of the broadness of this research field (www.bio.vu.nl/thb/deb/DEB_papers.pdf). The development of DEB theory started in 1979 with the work of Bas Kooijman, leading to a comprehensive first book on the theory in 1993; the third, and current, edition appeared in 2010 (Kooijman 2010). General overview papers have been published (Nisbet et al. 2000, Kooijman 2001, Jusup et al. 2017), as well as general papers on the application in ecotoxicology (Jager et al. 2006, Jager et al. 2014a).

[4] Jager, T. and E. I. Zimmer (2012): Simplified Dynamic Energy Budget model for analysing ecotoxicity data. Ecological Modelling 225: 74-81

[5] Jager, T. (2019). Making sense of chemical stress. Applications of Dynamic Energy Budget theory in ecotoxicology and stress ecology. Leanpub, https://leanpub.com/debtox_book; Version 2.0, 9 May 2019.

Figure 4: DEBtox – Schematic Diagram of the Energy Flows in a Standard DEB Animal



The nodes *b* and *p* denote switches at birth (start of feeding) and puberty (start of reproductive investment). The mobilisation flux is split into two branches according to a constant fraction κ. Graph reproduced from Jager (2019).

DEB theory constitutes a huge simplification of biology, based on an explicit set of assumptions. DEB is a theory for all life, but here the focus lies on the models for animals, and in particular the 'standard animal' model. The set of underlying assumptions can be summarised as (Jager 2019):

1. The basic state variables of the individual are reserve, structure and maturity. Reserve and structure have a different but constant composition (strong homeostasis), and maturity represents information (and therefore has no contribution to overall size or mass).
2. Life-stage events are linked to the level of maturity. When maturity exceeds a threshold value, the individual starts feeding ('birth'). A higher maturity threshold marks the start of reproductive investment ('puberty'). Above the puberty threshold, maturity does not increase any further.
3. Food is instantaneously converted into reserve, which in turn fuels all metabolic processes. The mobilisation rate of the reserve depends on the value of state variables of the individual only (reserve and body size).
4. The embryonic stage initially has a negligible amount of structure and maturity, but a substantial amount of reserve. The reserve density at birth equals that of the mother at egg formation (maternal effect). The costs for making a single egg thus depend on the reserve status of the mother. The developing embryo has the same value for the DEB parameters as the mother.
5. The maximum feeding rate is proportional to the surface area of the individual, and the food handling time is independent of food density. Assimilation efficiency is independent of body size or food density. The total time budget consists of the sequential processes of searching for food (which depends on food availability) and handling it (which does not).
6. At constant food density, the ratio between the amount of reserve and structure becomes constant (weak homeostasis). This assumption specifies the reserve dynamics, as a very specific function is needed to yield weak homeostasis.
7. Somatic maintenance is proportional to structural volume, but some components may depend on surface area (e.g., osmotic work in aquatic organisms and heating in mammals). Reserve does not require maintenance.
8. Maturity maintenance is proportional to the level of maturity.
9. A fixed fraction of the mobilised reserve is allocated to somatic maintenance plus growth, the rest to maturity maintenance plus maturation or reproduction (the κ-rule). Somatic maintenance takes priority over growth, and maturity maintenance takes priority over maturation (before puberty) or reproduction (after puberty). Maintenance costs are thus paid first, and what remains can be used for growth/reproduction.

10. The resources allocated towards growth are converted into structural biomass with a fixed conversion factor. The resources allocated towards reproduction are converted into eggs with a fixed conversion factor. Buffer-handling rules can be used to convert the continuous allocation to reproduction into discrete batches of eggs.
11. The individual does not change in shape during growth (isomorphism). This allows for a constant relationship between structural body volume and structural surface area. Changes in shape can be accommodated as extension of the standard model.

DEBtox (Jager and Zimmer 2012) is a simplified version of the standard DEB animal model; it is simplified to remove the dimension of 'energy' from the model system, and to reformulate the model into easy-to-interpret compound parameters (such as maximum body length and maximum reproduction rate) rather than the abstract primary parameters of DEB theory (such as specific maintenance costs). DEBtox is derived from the standard model by a set of additional assumptions (see Jager and Zimmer 2012):

1. The embryonic stage is not considered; the focus lies on standard ecotoxicological testing for growth and reproduction in juveniles/adults.
2. For juveniles, the maturity level is taken proportional to structure. Therefore, instead of a maturity threshold for puberty, we can take a threshold for structural body size (length at puberty). Thus, there is no need to follow maturity as a state variable. This assumption does not only have to hold for different food levels, but also under toxicant stress; size at puberty should thus not be affected by the toxicant.
3. The energetic costs for an egg are constant under all circumstances. This contrasts the assumption for 'maternal effects' in DEB theory, where the amount of reserve deposited into a single egg depends on the reserve status of the mother.
4. The reserve is always in steady state with the food level. This is, strictly speaking, only valid when we consider situations with constant food levels, or when the changes in food availability are slow relative to the dynamics of the reserve. However, a dynamic reserve can easily be included as additional state variable, without the need for additional parameters.

Other assumptions that are usually made when working with DEBtox models (although not absolutely required) is that there is no reproduction buffer (offspring production is viewed as a continuous process), and that the measured body size of the organisms is proportional to the structural size as used by the DEB model. Apart from assumptions about the animal's life cycle, DEBtox also requires assumptions for the effects of chemicals on the individual:

1. A chemical first needs to be taken up in the body before it can exert a toxic effect. DEBtox applies a one-compartment model with first-order kinetics, including growth dilution of the internal concentration, and scaling of the rate constants with the surface:volume ratio of the organism. Note that more complex TK models can easily be included, but parameterising them rapidly requires more data than available in standard ecotoxicity tests.
2. The internal concentration affects one or more metabolic processes in the animal, typically focussing on assimilation, maintenance, growth, reproduction costs, or direct effects on embryonic development. The affected process is called the metabolic or physiological mode of action (pMoA, Alda Álvarez et al. 2006a, Ashauer and Jager 2018), and is generally determined by fitting the model with various pMoAs to the data and selecting the best-fitting one.
3. The relationship between the internal concentration and the affected model parameter, governing a specific process, is taken as linear with a threshold (the no-effect concentration). This is the same relational form that is used in GUTS. However, at the moment, the concept of 'damage' (which has a central position in GUTS) has not explicitly been considered in DEBtox models until very recently (Jager, 2020).

Applications and Data Requirements

DEBtox models (and related ones) have been used for a wide range of applications (see list of papers at http://www.debtox.info/papers_debtox.html), but the typical use is to interpret toxicity data for the endpoints growth and reproduction over time. Advantage over the use of dose-response curves is that all data over time are used, that growth and reproduction data can be analysed together (as these endpoints are metabolically linked), that a true no-effect concentration can be estimated with a confidence interval, and that specific problems with the data set can be identified. As an example of the latter: food limitation for a part of the test leads to specific deviations from the expected patterns (Zimmer et al. 2012) that may prompt requests for further testing. Furthermore, there is no need for the conditions to be constant during the test (see e.g., Pieters et al. 2006, Billoir et al. 2011), and as the model has a mechanistic basis, predictions can be made for untested conditions (e.g., to extrapolate from constant exposure to time-varying exposure scenarios, or ad libitum food availability to food limitation). For these reasons, DEBtox models are also used as individual-level module in population effects modelling (see review in Jager et al. 2014a).

Typically, DEBtox models are fitted to data for growth and reproduction over time, at different exposure concentrations, simultaneously. DEBtox models can be linked to a GUTS model to include effects on survival over time as well (Jager et al. 2004, Jager and Zimmer 2012). We seldom know the pMoA beforehand. However, the different pMoA's yield different patterns of effects on growth and reproduction over time; the best-fitting pMoA is taken as the more representative one. In some cases, different pMoA's provide a similar fit to the data, and it is impossible to select a single one as the best one (this is particularly common with effects on assimilation and maintenance). For the estimation of the no-effect concentration, this is generally not so important (Kooijman and Bedaux 1996b), but it may well affect the extrapolation to the population level (Martin et al. 2014).

DEBtox was initially specifically developed to allow for the analysis of standard ecotoxicity tests, in particular the 28-day fish growth test and the 21-day Daphnia reproduction test. However, these tests are certainly not optimal for modelling. The type of data sets that are most useful for DEBtox analysis are those where growth and reproduction are followed over a good part of the life cycle (starting before the first brood, and stopping when body size is close to the asymptotic maximum size and several reproductive events have been captured). The fish-growth test protocol only prescribes size measurements at day 0 and day 28 (with the possibility of an additional measurement at day 14). This is very little information to estimate all model parameters, although defaults can be used for the biological parameters (Kooijman and Bedaux 1996c). A more structural problem is that the fish growth test can only be used to identify the pMoA's that affect growth directly or indirectly: assimilation, maintenance or growth costs. Some chemicals may affect reproduction directly, and this can happen at much lower concentrations than those that affect growth (Jager et al. 2007, Jager and Zimmer 2012). Specific effects on reproduction cannot be identified from effects on growth alone, which is also a problem for the current dose-response analyses of such data. This limits the usefulness of the fish growth test. The Daphnia reproduction test is a far more useful test for DEBtox modelling, as it covers a considerable part of the life cycle (starting with neonates, which are followed into adulthood, with several reproductive events). However, the standard test protocol does not prescribe measurement of body size (although it is advised). Measuring body size (preferably at several time points) drastically increases the possibility of identifying the pMoA from the data. Regarding some other standard tests, the earthworm reproduction test is not useful, as the test starts with adult worms, and reproduction (and body weight) are only assessed at the end of the test. Similarly, the standard test with springtails is of little use. The fish two-generation test may be useful when body size is determined regularly as well. However, so far, such tests do not seem to have been analysed with DEBtox models yet.

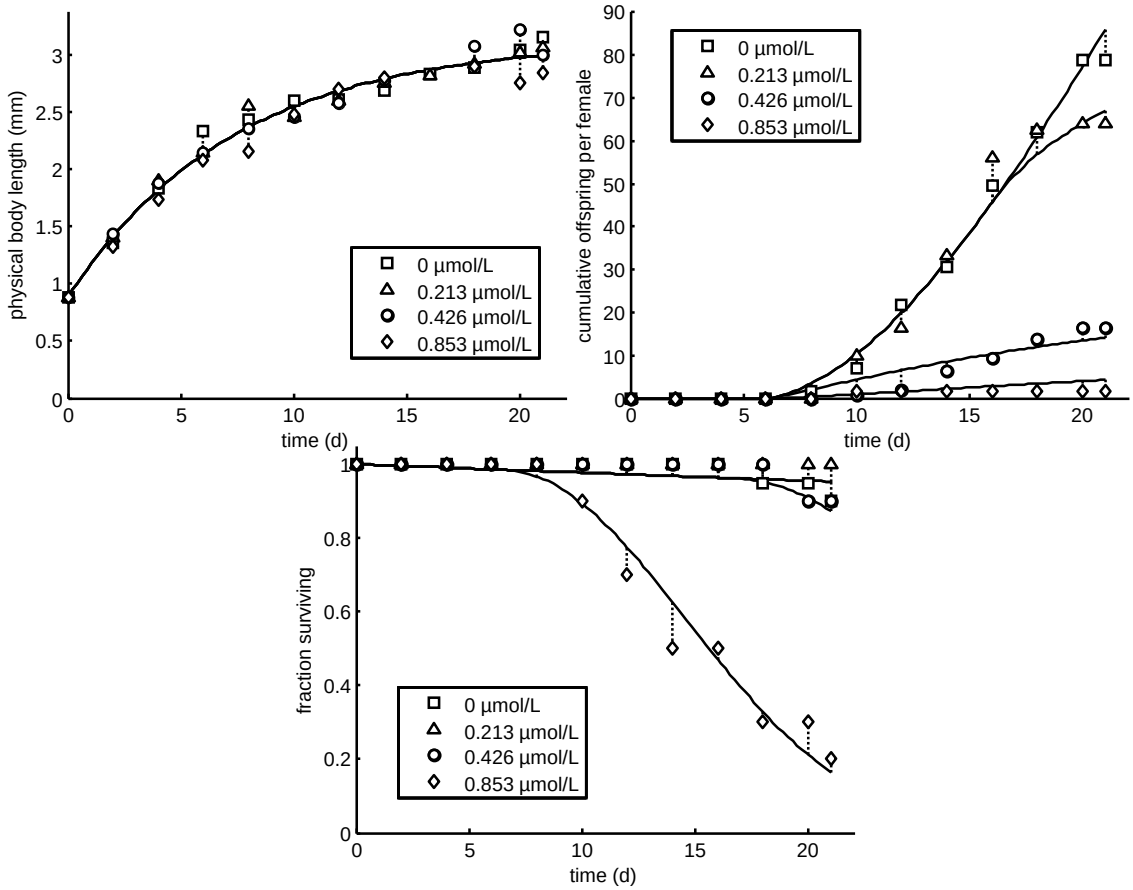
It is important to stress that there are no minimum data requirements for using DEBtox. Data sets with little information simply lead to large confidence intervals on the model parameters, and may not allow for a meaningful identification of the parameter and/or the pMoA. Data limitations may be addressed by setting default values (or Bayesian prior distributions) for certain parameters. In the original DEBtox software, Daphnia reproduction data were analysed in absence of any information on growth; for the parameters governing growth, typical defaults for Daphnia were set (Kooijman and Bedaux 1996b). If DEBtox models are to be used for ERA, it makes sense to develop/adapt test protocols to make the tests more suitable for the analysis with DEBtox (see e.g., Barsi et al. 2014). Current test protocols were developed with the requirements of classic dose-response analysis in mind, which are very different from those needed for mechanistic modelling.

Example Fit

As an example, a case study for fluoranthene in *Daphnia magna* is provided here, as used in Jager (2019). Figure 2 shows the simultaneous fit to body length, cumulative reproduction, and survival, over time. Parameter estimates with their confidence intervals are provided in Table 1. The data used are from Jager et al. (2010). The parameter g governs the influence of the reserve compartment on the model curves. This parameter cannot usually be estimated from such data sets (g is a very insensitive parameter in most cases), and in this case that is also not possible. Therefore, this parameter is fixed to a default value, as done in most DEBtox applications. Also fixed are the initial body length of the Daphnia, and the food level (f=1 implies ad libitum food availability). The remaining 10 parameters are fitted to the data for the three endpoints simultaneously. Given the available data (the points in Figure 2), the number of fitted parameters is very reasonable. The mode of action that was used was an increase in the costs for reproduction, which provided the best fit to the data.

Interestingly, the parameters governing the toxicity are poorly defined by the data; the confidence intervals are extending to zero or to infinity for all parameters. The origin of this problem lies with the elimination rate k_e. This parameter is not well identified from the data, and a very low elimination rate also fits the data well. A very low elimination rate implies linear uptake of the chemical into the body over time. This in turn implies that even very low external concentrations could lead to effects over very long periods of exposure, and hence the confidence intervals of the two no-effect concentrations (one for lethality and one for sub-lethal effects) extend to zero. It should be noted that poor identifiability of parameters does not necessarily translate into uncertain predictions of effects, due to the possibility of strong correlations between parameters (as will also be the case in this example). The joint uncertainty in model parameters can still be translated into useful model predictions, for example in the form of a population growth rate with confidence intervals (Jager and Zimmer 2012).

Figure 5: DEBtox – Example Fit to *Daphnia magna*



Simultaneous fit to body length, cumulative offspring produced, and survival using the DEBtox model. The data are for fluoranthene in *Daphnia magna*. Graphs reproduced from Jager (2019).

Table 7: DEBtox – Example Parameter Estimates for *Daphnia magna*

Symbol	Parameter	Value (95% CI)	Unit
g	Energy-investment ratio	1 (n.e.)	(-)
L_0	Initial body length	0.88 (n.e.)	mm
L_p	Length at puberty	2.2 (2.1-2.2)	mm
L_m	Maximum body length	3.1 (3.1-3.2)	mm
r_B	Von Bertalanffy growth rate	0.14 (0.13-0.15)	d^{-1}
R_m	Maximum reproduction rate	10 (9.3-11)	offspr. d^{-1}
f	Scaled food availability	1 (n.e.)	(-)
h_0	Background hazard rate	2.3 (0.39-7.2) 10^{-3}	d^{-1}
k_e	Elimination rate constant	0.018 (<0.091)	d^{-1}
c_0	No-effect concentration for sub-lethal effects	0.038 (<0.13)	µM
c_T	Tolerance concentration	0.0050 (<0.026)	µM
c_{0s}	No-effect concentration for lethal effects	0.079 (<0.32)	µM
b	Killing rate for survival	2.2 (>0.32)	$\mu M^{-1}\ d^{-1}$

Parameter estimates for the fit in Figure 5 with 95% likelihood-based confidence intervals. Data from Jager (2019).

Related DEB-based Models

DEB is a theory with almost 40 years of history and a broad community of users. Many groups have been working on DEB in the context of ecotoxicological effects, which has led to a range of closely-related models. For this evaluation, the focus lies on the set of DEBtox equations presented by Jager and Zimmer (2012). The original publication of the model for reproduction data (Kooijman and Bedaux 1996b) contained a few derivation errors in several of the equations (see Billoir et al. 2008b, Jager and Zimmer 2012). The set of equations by Billoir et al. (2008b) corrected most of the errors in the original equations, but introduced a new one. This is of minor consequence, affects the growth rate, and only shows up when food level in the model is limiting (f<1) (see Jager and Zimmer 2012). These errors all resulted from the difficulties in deriving the simplified DEBtox equations from the standard DEB equations. DEBtox applies various scaling (to remove the dimension of 'energy' and to work with easy-to-understand compound parameters such as maximum length and maximum reproduction rate). Applying these scales in a consistent manner, when parameters can also be affected by toxicant stress, turns out to be a non-trivial matter. For this reason, Jager and Zimmer (2012) reformulated the model using unscaled body length, provided an extensive derivation (so that it can be checked by others), and applied the stress factors on basic model parameters, rather than incorporating them into the (rather tortuous) equations for growth and reproduction in terms of compound parameters. These steps increase transparency and reduce the chances of errors in the derivation or the application of toxic stress.

Muller et al. (2010) present a version that stays closer to the formulation of the standard DEB animal model, using mainly the primary parameters (parameters that directly relate to metabolic processes) rather than easy-to-interpret compound parameters (combination of two or more primary parameters). These authors make some simplifying assumptions regarding the TK, and only consider two modes of action: decrease in assimilation and increase in maintenance (and the combination of these two).

The complete standard DEB animal model has also been used to analyse toxicity data. Examples can be found elsewhere (Jager and Klok 2010, Jager and Selck 2011, Augustine et al. 2012). The full model does not lead to a different interpretation of the toxicity data, as was demonstrated by Jager and Klok (2010), and by comparing the analysis in Jager and Selck (2011) with the case study in Jager (2019). However, working with the standard DEB animal model has the advantage that model parameters for a species can be taken from the extensive AmP library of DEB parameters (http://www.debtheory.org/wiki/index.php?title=Add-my-pet_Introduction) (Marques et al. 2018), which is maintained by a group of 'curators' (who also evaluate new entries). Disadvantage is that the standard model is relatively complex (compared to DEBtox and DEBkiss models) and requires considerable expertise to fit. An EFSA-commissioned project has recently investigated the possibilities for using the standard DEB model for risk assessment in food and feed safety, with a focus on mixture toxicity and effects at the population level (http://deb.akvaplan.com/efsa.html) (Baas et al. 2018). Zimmer et al. (2018) recently published a case study for application of standard DEB, using AmP for the basic parameters, to analyse early-life stage toxicity in rainbow trout (that case study is evaluated in detail in section 3.3).

It is good to realise that the add-my-pet library is not aimed at regulatory application, but rather to present the best-possible parameter set, at this moment, for a species. For some species, the entry is based on a large amount of (good-quality) data, and there can be a large degree of confidence in the appropriateness of the parameter values. However, for others, either less data is available, or a less-thorough data search was performed. In such cases, the data are supplemented by 'pseudo-data' (based on theoretical relationships with maximum body size) in the calibration procedure to arrive at a best guess. Therefore, the quality of the entries varies considerably (and entries may be updated at any moment). Whether that matters or not depends on the particular application that the model is

used for. In general, the quality of the basic parameter set will become more important when the data from the toxicity test contain little information and with increasing extrapolation distance.

A more recent member of the DEB family is DEBkiss (Jager et al. 2013, Jager 2018). This model framework is simplified from the DEB animal model by completely removing the reserve compartment, and by simplifying the embryonic stage. Removing the reserve is consistent with the fact that the DEBtox parameter governing the size of the reserve (g) turns out to be very insensitive in most model fits. The result is a simple and transparent set of model equations, which are much easier to communicate, understand and apply than the DEB (or even DEBtox) formulations. Originally intended for educational purposes, DEBkiss turns out to be highly practical for many applications (see http://www.debtox.info/debkiss_appl.html). In many cases, especially for the small invertebrates that most commonly feature in ecotoxicology, removing the reserve compartment has no major impacts on model behaviour (see e.g., Jager and Klok 2010, Martin et al. 2013a). DEBkiss is very similar to the original DEB model of Kooijman and Metz (1984a), which was applied in ecotoxicology by Klok and coworkers in a range of papers (o.a., Klok and De Roos 1996, Klok 2008). The e-book on DEBkiss (Jager 2018) provides proposals to update the 'DEBtox model' for applications in ecotoxicology (a.o., basing the model on DEBkiss and explicitly considering 'damage'). This was further worked out in a DEBkiss-based revision of DEBtox, which was published by Jager (2020). This revision brings DEBtox in line with the TKTD formulation established for GUTS, and includes a starvation module (which is needed to consistently deal with time-varying stress on assimilation or maintenance). This evaluation does not specifically consider this most recent revision, but focusses on the previous DEBtox version of Jager and Zimmer (2012).

DEBkiss can be applied in the form of the full model, with explicit mass balance, and direct access to metabolic processes, which makes it easier to expand (e.g., to include other endpoints, such as respiration and feeding), and allows it to be applied to non-feeding early-life stages. However, it is also possible to formulate it with easy-to-interpret compound parameters (Jager 2018). This yields a set of equations that is mathematically equivalent to those of DEBtox, with g set infinitely large (whereby reserve drops out). Its simplifications make DEBkiss suitable for many practical applications where simplicity and transparency are key, but removing the reserve makes the DEBkiss analyses difficult to compare to the substantial work done with the standard DEB model. For example, the extensive add-my-pet library of DEB parameters cannot be used for DEBkiss purposes. On the other hand, this model can usually be fitted to experimental data without requiring this data base.

All these model versions are very closely related, and will draw similar conclusions from a set of data. However, especially in more extreme extrapolations, differences in the predictions may occur. This range in model versions reflects the large and diverse user community of DEB-based models, as well as the diversity in applications.

2.3.1.2 **Status of the Model**

The original DEBtox models for sub-lethal effects (Kooijman and Bedaux 1996c, Kooijman and Bedaux 1996b) have been included in OECD/ISO guidance on the statistical treatment of ecotoxicity data (ISO 2006, OECD 2006) under the header 'biology-based methods'. This work is also mentioned in REACH guidance (ECHA 2008). However, few DEBtox analyses have been submitted as part of risk assessment dossiers (an example is Zimmer et al. 2018, which concerns an atypical application to early-life stage toxicity). DEBtox is one of the models treated in the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018), where the following conclusion was reached: "Based on the current state of the art (e.g. lack of documented and evaluated examples), the DEBtox modelling approach is currently limited to research applications. However, its great potential for future use in prospective ERA for pesticides is recognised."

DEBtox can be used in ERA in different ways. The most prominent ones are:

1. Analysis of data from toxicity tests, using all of the data from the test over time. The no-effect threshold from DEBtox can be used as summary statistic, or DEBtox can be used to calculate EC*x*,*t*, for any effect level *x* and any exposure time *t*, for both growth and reproduction. Furthermore, the analysis can reveal inconsistencies in the experimental data, or between data sets, and thus be used as a quality check on the data.
2. Extrapolation of effects to untested environmental conditions. I.e., predicting growth and reproduction over time for a specific exposure scenario.
3. As building block in population/community models.
4. Extrapolation between chemicals and species. There has been quite some study into the patterns of basic DEB parameters between species. However, there is very little work to show whether the pMoA and the toxicity parameters are the same for the same chemical in different species (or whether there are clear patterns). This issue, and a way forward, is discussed in a recent paper (Ashauer and Jager 2018). Therefore, the possibilities for such extrapolation are currently limited.

Application 1 is the most common one, and also the one that the OECD/ISO guidance (ISO 2006, OECD 2006) focusses on. Application 2 is used, in addition to Application 1, in several papers (e.g., Jager et al. 2004, Alda Álvarez et al. 2005, Alda Álvarez et al. 2006a, Jager and Klok 2010), as part of the prediction of population growth rates for conditions of limited food availability or different environmental temperature. Application 3 is reviewed in (Jager et al. 2014a), and Application 4 is rare and requires further study.

From the list of potential applications, the recent EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on application 2 and proposes the same workflow as for GUTS models: 1) calibrate the DEBtox model to toxicity data for constant exposure, 2) validate the calibrated model using additional toxicity data for pulsed exposure, and 3) derive profile-specific EP*x* values (factor by which an exposure profile from a fate model must be multiplied to reach *x*% effect at the end of the profile). However, no examples of such an extrapolation are provided for DEBtox, and several issues would need to be settled. For Daphnia, for example, it would make little sense to start with a neonate, and follow its cumulative reproduction for a >1 year exposure scenario.

Several software platforms exist that can perform DEBtox calculations.

- The standalone Windows software 'DEBtox' with its documentation (Kooijman and Bedaux 1996a) is no longer offered or maintained. It had a simple user interface, but was limited to several standard tests only (fish growth, Daphnia repro, algae population growth, and acute survival, all at constant exposure). As the software was limited to fitting standard tests, data analysis was simply a matter of entering the data and pushing a button.
- Matlab packages for DEBtox and DEBkiss are offered for free download, as part of the BYOM platform (maintained by Tjalling Jager, DEBtox Research, The Netherlands), which is very flexible but with limited user-friendliness and limited user manual: http://www.debtox.info/byom.html. Working with this package requires a basic working knowledge of Matlab, expertise with model optimisation, and knowledge of the DEBtox concepts (this is not push-button software). Note: the DEBtox BYOM package is the only software to date that implements the DEBtox version as presented by Jager and Zimmer (2012).
- Other researchers have developed their own DEBtox versions in R (Billoir et al. 2008b, Goussen et al. 2015) and WinBugs (Billoir et al. 2008a). However, these are not publicly available.
- The standard DEB model is part of the freely-downloadable Matlab platform DEBtool: http://www.bio.vu.nl/thb/deb/deblab/index.html. User-friendliness is currently low, but a considerable group of researchers is working with this software, and further developing it. This is also the software that is used for the add-my-pet library, and in the international DEB course, see: www.debtheory.org/wiki/.

An R package for standard DEB and toxicant stress effects has been under development as part of an EFSA-funded project (http://deb.akvaplan.com/efsa.html). This is expected to be released for free download via EFSA at some point (including standard R-package documentation). However, that package only predicts effects, given a certain parameter set (it does not fit toxicity data).

This model evaluation will refer to DEBtox as the simplified model in the formulation published by Jager and Zimmer (2012), unless noted otherwise.

2.3.2 Model Description

2.3.2.1 Problem Definition

Context in which the model will be used

DEBtox is a model for animal bioenergetics. It deals with effects on life-history traits, in particular growth and reproduction (which are usually included in chronic toxicity testing). DEBtox is a model for an individual animal, as it develops over time, in principle from birth (start of feeding) to death (the over-arching DEB theory also deals with embryonic stages, and non-animal life). DEBtox is not restricted to a particular tier; it can be used in Tier 1 to analyse data from toxicity tests with sub-lethal endpoints, but also in higher tiers to analyse more complex non-standard data sets (e.g., with time-varying exposure), to predict traits for untested exposure conditions, or as module in Higher Tier models (see list of possible applications in the general text above). DEBtox is often linked to (a specific case of) GUTS to include survival aspects as well. The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on potential application of DEBtox in Tier 2C.

Specification of the question(s) that should be answered with the model

Questions that relate to the growth and reproduction of individuals, as function of time and exposure profile (the profile of concentration versus time, e.g., output from fate models).

Specification of necessary model outputs and protection goals

The model can be used to analyse (standard) toxicity data, estimate an ECx,t (for any effect percentage x, and any time point t), and predict growth and reproduction as function of time for any exposure profile. The model is thus relevant for protection goals that deal with individual performance (e.g., vertebrates), as well as cases where individual-level effects are an important aspect of the population impacts (e.g., in combination with a population model).

Domain of applicability of the model

Analysing and predicting growth and reproduction of individuals. DEBtox can be used for many, in principle all, animal species and all chemicals (as well as many non-chemical stresses). However, several species will require model adaptations to accommodate their life cycles; examples are insects and copepods, which do not grow after a final moult. The over-arching DEB theory deals with all forms of life, and could therefore in principle be used to analyse/predict toxicant effects on other (non-animal) species as well. The applicability domain is mainly determined by the available data for calibration and testing of the model (validation), and the extent of extrapolation (uncertainty increases with increasing distance between the situation for calibration and the situation for prediction).

Why is the model being used?

Because standard approaches for analysis of toxicity data (e.g., NOEC or dose response curves to estimate EC50) do not consider exposure time and exposure profile, and thus introduce bias in the risk assessment (Jager 2011), cannot accommodate non-standard toxicity data (e.g., when exposure is not constant), and cannot be used for meaningful predictions (e.g., for time-varying exposure profiles in the field).

What protection goal is being addressed?

Protection goals that relate to the growth and reproduction of individuals. DEBtox can also be used as individual-level module in models at higher levels of organisation (e.g., population models), and thereby aid in addressing other protection goals.

What outputs are required?

The model can be used to analyse (standard) toxicity data, such as those resulting from the Daphnia reproduction test. Output are model parameters (including a time-independent threshold for effects) with confidence intervals. These model parameters can subsequently be used to estimate an EC*x*,*t* (for any effect percentage *x*, and any time point *t*), and predict growth and reproduction as function of time for any exposure profile or set of environmental conditions.

How was the species chosen?

DEBtox can, in principle, be used for any species (though so far the focus has been on animals).

Which other species/groups are being covered by the chosen one(s)?

DEBtox can, in principle, be used for any species. If the model is calibrated to data for one species, the parameterisation reflects that species (and perhaps even the life stage that was tested). This is similar to the limitations of a dose-response curve (or EC50 derived from it), with the remark that DEBtox explicitly deals with effects over time and is thus better suited to identify the mechanistic basis underlying sensitivity differences. There is currently insufficient information to identify patterns in the parameters governing the toxic response across species (see Ashauer and Jager 2018). However, a lot of work has been done on the DEB(tox) parameters governing the unstressed life history of animals, and on how they vary across species (see Marques et al. 2018, and references therein).

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

That depends on the specific application: it depends on the data that are being used and the question that is to be answered. For example, if DEBtox is to be used to predict growth and reproduction for an untested exposure profile, it would be useful to test the predictions of the calibrated model with a few additional toxicity tests (with another exposure profile). Validation with pulse-exposure toxicity tests is therefore required for aquatic risk assessment under the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018), for each case. The opinion proposes several goodness-of-fit measures, but no explicit pass-fail criteria are provided. However, even if the model is not tested for a specific application, a DEBtox prediction would still constitute a better-educated guess than predictions from an EC50 or NOEC (which is also a model, albeit a very poor one).

2.3.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

DEBtox itself is not built on data and does not include data in any way; it is a generic model for which all of the parameters are species (and chemical-) specific. However, for a specific application, the model would need to be parameterised using data for the species and chemical of interest. When using simplified DEBtox models (which is the focus of this evaluation), the model can be fully parameterised using the toxicity data for a specific chemical-species combination. When using the full standard DEB model, or accommodating limited data availability (as is e.g., the case for ELS studies), it will generally

not be possible to fit all model parameters on the data from toxicity tests alone. In such cases, the basic (non-chemical-specific) parameters of the DEB model could be taken from the add-my-pet library (http://www.debtheory.org/wiki/index.php?title=Add-my-pet_Introduction). However, the limitations of this library should be considered (see general text above, and comments under 'data quality' below).

The data that are available for calibration and testing depends on the specific application. In general, for calibration, DEBtox requires data for growth and reproduction of individuals over time, for a considerable part of their life cycle. The exposure concentration does not need to be kept constant, and, in fact, if the model is to be used to predict effects due to time-varying exposure, it makes sense to calibrate the model using data from tests with non-constant exposure (to minimise the distance for extrapolation). The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) proposes a range of requirements to data sets for calibration and validation for the purpose of aquatic ERA of PPPs (for validation: toxicity tests run under pulsed exposure, Section 4.1.4.5).

Assessment of the quality of the data

That depends on the specific application and the specific species-compound combination that is considered; the model itself does not contain any case-independent data in any way. The currently-used quality controls on toxicity testing also apply here (e.g., criteria laid down in test guidelines). However, these criteria are not necessarily all relevant for DEBtox application (e.g., the demand for constant exposure in standard tests). The EFSA opinion discusses the relevance of the validity criteria for OECD aquatic studies for TKTD modelling (Table 6), and proposed criteria for proper validation tests. The criteria should be applied for each model analysis.

Data quality criteria should also depend on the specific DEB model that is used. The classical DEBtox is able to work with the results from some standard toxicity tests (e.g., the Daphnia reproduction test, if body size is determined, preferably at a few time points), as well as non-standard tests (e.g., there is no stringent need for constant exposure). Use of the standard DEB model requires the basic parameters from add-my-pet (see above), which has its own quality control: each entry has been approved by a board of curators, and is accompanied by marks for completeness of the data set and the goodness-of-fit. However, this check is mostly technical (whether the code is used in the right way and everything is well documented), and does not relate to the requirements of risk assessment per se (e.g., a poor fit or limited data availability does not preclude a parameter set being entered into the collection).

2.3.2.3 Conceptual Model

Description of the model concepts including a diagram

See general text above, and flow diagram already presented in section 2.3.1. In short, reproduction and growth need to be fuelled from feeding (in animals). A reduction in growth and reproduction in a stressed individual thus implies a change in its energy budget: either less food was taken in or it was used in a different way. DEBtox is based on the well-established energy-budget rules of DEB theory, which can be illustrated by the scheme below. DEBtox is simplified from the DEB scheme as the reserve dynamics are generally assumed to be very fast, maturity is not explicitly considered (reproduction starts at a fixed body size), and neither is the reproduction buffer (reproduction is continuous).

The energy that is assimilated from food is placed in the reserve, from which it is mobilised to fuel the energy-requiring metabolic processes. The mobilisation flux is split according to a fixed fraction of kappa to the soma (somatic maintenance and growth), and 1-kappa to maturation/reproduction. Maintenance costs are paid first on each side, and the remainder is used for the subsequent processes.

Identify the main components and processes in the system

DEBtox consists of a toxicokinetics module and a simplified DEB model.

How the effects of the chemicals are modelled

To include toxicokinetics, a one-compartment model is used. This is basically the same scaled model that is used in GUTS, generally extended with the effects of changes in size (growth dilution and changes in the surface:volume ratio). The internal toxicant above a threshold (the no-effect concentration) changes one (or several) of the energy fluxes in the individual in a linear manner. The specific flux that is affected is chemical-specific and is called the physiological mode of action (pMoA). In the standard DEBtox model, five pMoA's are considered.

How the components and processes are linked

See explanation above.

2.3.2.4 Formal Model

Identification of the model variables

In the DEBtox model as presented by Jager and Zimmer (2012), the state variables are the (scaled) internal concentration, the structural body size of the individual, and the cumulative reproduction rate. Reserve dynamics can easily be added if required (adding the reserve as an extra state variable, but without the need for additional parameters). Note that the formulation for the standard DEB model is different, and is not treated in this evaluation.

Identification of the model parameters

g	Energy-investment ratio (this parameter is generally set to a default)
L_0	Initial body length
L_p	Length at puberty
L_m	Maximum body length
r_B	Von Bertalanffy growth rate
R_m	Maximum reproduction rate
f	Scaled food availability (f=1 means ad libitum)
k_e	Elimination rate constant
c_0	No-effect concentration for sub-lethal effects
c_T	Tolerance concentration

Description of the most important model equations or algorithms

Reserve dynamics (can be ignored by setting *e*=*f*):

$$\frac{d}{dt}e = (f - e)\frac{v}{L} \qquad \text{with } e(0) = 1$$

Change in body length:

$$\frac{d}{dt}L = \frac{k_{Mg}}{3(e+g)}\left(e\frac{v}{k_{Mg}} - L\right) \qquad \text{with } L(0) = L_0$$

Reproduction rate (which will be cumulated over time):

$$R = \begin{cases} 0 & \text{if } L < L_p \\ \frac{R_{m0}}{L_{m0}^3 - L_p^3}\left(\left(\frac{v}{k_M}L^2 + L^3\right)\frac{e}{e+g} - L_p^3\right) & \text{otherwise} \end{cases}$$

The compound parameters k_M and v can be calculated from more easy-to interpret parameters r_B and L_m, and g. Note: the zero in the subscript indicates that this is the parameter's value in the control. The value of the parameters can change due to stress (see stress factor s below), depending on the selected pMoA.

$$k_{M0} = r_{B0}\frac{3(1+g_0)}{g0} \qquad \text{and} \qquad v = L_{m0}k_{Mo}g_0$$

Change in scaled internal concentration over time (the k_e is referenced to the value in a fully-grown adult in the control):

$$\frac{d}{dt}c_V = k_e^{ref}\frac{L_{m0}}{L}(c_d - c_V) - c_V\frac{3}{L}\frac{d}{dt}L$$

This model is essentially the same as the scaled damage model used in the reduced GUTS models, extended with effects of growth (growth dilution and linking the elimination rate to the surface:volume ratio of the individual). The internal concentration is converted into a stress factor. The target parameter in the energy budget (the pMoA) will be multiplied or divided by this factor:

$$s = \frac{1}{c_T}\max(0, c_V - c_0)$$

2.3.2.5 Computer Model

Description of the model implementation

There are several model implementations available for various formulations of DEBtox, probably around 10. However, there is no user-friendly standalone executable anymore. See also the list of model implementations in section 2.3.1.

Checking the computer model for errors, bugs and inconsistencies in the code

Each implementation has its own development history. Several implementations have been extensively used and tested, although these tests have seldom been formalised and documented at this point.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

Each implementation has its own development history. Several implementations have been extensively used and tested, although these tests have seldom been formalised and documented at this point.

2.3.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context in which the model is run

For DEBtox, the scenario entails the exposure profile (i.e., the exposure concentration as function of time), and the conditions in the toxicity test (e.g., temperature and food availability). This scenario should thus be clear for each calibration analysis (and match the design of the toxicity test) and for

each extrapolation (e.g., the exposure profile from a fate model). DEB-theory provides rules for how temperature and food availability affect the life history of the organism. However, also the toxicity-related parameters may be affected by such environmental conditions (in general, these are treated as constant).

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

The values of all DEBtox parameters (with the exception of *g*, for which a default is generally set) are established by fitting the model to a toxicity data set (only for using the standard DEB model, the add-my-pet library would be needed). It is good to realise that when DEBtox is used for predicting effects under time-varying exposure profiles, the pesticide concentrations will have been generated using the relevant exposure models. These exposure models will include considerations of environmental factors such as soil type, rainfall and agronomic practice.

2.3.2.7 Parameter Estimation

Description of the model parameter estimation

The values of the DEBtox parameters are established by fitting the model to a toxicity data set (reproductive output and body size over time in various treatments). DEBtox can be calibrated using the data from several standard test protocols (notably the Daphnia reproduction test), as long as observations on reproduction (and preferably body size) are reported at several points in time. However, it can also be calibrated on non-standard data sets, for example, data sets where the exposure concentration was not kept constant over time. It should be noted that standard test protocols have not been optimised for the purpose of calibrating mechanistic models; they were optimised for fitting a dose-response curve at the end of the test to derive an EC*x* or NOEC. Other test designs will be far more efficient for the modelling purpose. Optimal test design will depend on the specific properties of the test chemical, as well as on the way that the model is applied (see also Barsi et al. 2014). In general, useful toxicity tests will include observations on body size and reproductive output over time, over a good part of the life cycle (starting with juveniles and continuing long enough to allow estimation of maximum body size and the reproduction rate).

Parameters estimated from the literature — what are the sources and why are these appropriate?

For the simple DEBtox models, no parameters are estimated from the literature, unless raw toxicity data are extracted from publications. The parameter *g* could be set using the add-my-pet library, although this parameter is generally not sensitive within a realistic range of values (generally, a default is set). Using the standard DEB model implies that basic parameters from add-my-pet need to be used: each entry contains the references to the data sources (all open literature or data bases).

Parameters obtained from calibration — how and why this was done?

The values of the DEBtox parameters are established by fitting the model to a toxicity data set. This is done by maximising a likelihood function based on the normal distribution. Different software implementations use different numerical schemes for optimisation (e.g., Bayesian inferences or Nelder-Mead simplex optimisation).

2.3.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

Sensitivity analysis is pointless for models that are parameterised by fitting them to data; this holds for DEBtox as well as for dose-response curves (Jager and Ashauer 2018a). The information on sensitivity and identifiability is contained in the (joint) confidence interval of the parameters as a result of the calibration. If required, a classical sensitivity analysis *can* be performed, but it would have to be done after the model is calibrated to a specific data set (and the results will be uninformative).

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

In classical uncertainty analysis, all parameters receive an (independent) distribution, which is propagated to the model output. This is done once (or several times during development) for the model by the developer and serves as a 'mark of model quality' (EFSA PPR 2014b). This is also a rather pointless exercise for models that are completely parameterised by fitting them to data (Jager and Ashauer 2018a): we would need to give all parameters a distribution between zero and infinity as they are completely dependent on the data set and thus case-specific. The result of such an analysis would be meaningless. Instead, the uncertainty in the parameters should be taken from the model fit, which implies that an uncertainty analysis has to be done in each case, for each fit. In general, such a procedure is not called an uncertainty analysis but referred to as error propagation. Uncertainty in the model parameters can be quantified and propagated to uncertainty in the model predictions using various methods (Bayesian or frequentist based). This is a very important analysis as it shows the impact of the uncertainties in the parameter identification (even though uncertainty in the exposure profile for the predictions will also play a role, which is generally ignored). An example can be found in Jager and Zimmer (2012).

2.3.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

This item cannot be addressed for the model in general; the model output can only be compared to independent data after the model has been calibrated using data for a specific species-chemical combination. Such comparisons thus can only be performed as part of a specific application (dossier), where they can increase confidence in the model and its parameterisation for the specific case at hand. For models that are fitted to data, comparison to independent data only makes sense when the model is used in extrapolation; e.g., when DEBtox is calibrated to data for constant exposure and used to predict survival due to pulsed exposure. Such studies are rare, as usually, all data are used in calibration. An example of comparison to independent data is the prediction of mixture effects from single chemicals (Jager et al. 2014b). However, there seem to be no published examples of extrapolation from constant to time-varying conditions at the moment.

The workflow laid down in the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) explicitly specifies a validation step with pulsed-exposure testing for each species-chemical combination.

Demonstration that the model output provides an adequate match to data patterns

DEB-based models have been used for many decades and generally provide a good fit to growth and reproduction data (with and without stress) over time. A list of publications is maintained at http://www.debtox.info/papers_debtox.html (currently, around 90).

2.3.2.10 **Model Use**

Explanation of how the model conforms to the requirements set in the problem definition

This item cannot be addressed in general. DEBtox should be seen as dose-response model, but then way more robust and more mechanistic than descriptive methods (such as fitting a log-logistic dose-response curve at one time point, or hypothesis testing to derive a NOEC).

Description how the model works (user manual).

A full model description is available in the open literature (most complete one in Jager and Zimmer 2012) and a detailed e-book with technical annex (Jager 2019).

Description of the pesticide parameters values used in the model

The values of DEBtox parameters are established by fitting the model to a toxicity data set for the specific pesticide. The model itself contains no parameter values whatsoever.

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed in general.

2.3.3 Model Evaluation

Note that this section has been filled according to Appendix B, Summary checklist for model evaluation by the risk assessor, proposed in the EFSA Sci. Op. on GMP (EFSA PPR 2014b). Since then, the recommended checklist has been adapted to DEBtox models (EFSA PPR 2018).

2.3.3.1 Problem Definition

The regulatory context in which the model is run

This cannot be answered for the model in general. DEBtox deals with effects on life-history traits, in particular growth and reproduction. It is not restricted to a particular tier; it can be used in Tier 1 to analyse data from toxicity tests with sub-lethal endpoints, but also in higher tiers to analyse more complex non-standard data sets (e.g., with time-varying exposure), to predict traits for untested exposure conditions, or as module in Higher Tier models (see list of possible applications in the general text above). The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on application of DEBtox in Tier 2C.

The question that has to be answered with the model

This cannot be answered for the model in general. DEBtox is useful for questions that involve effects on growth and reproduction over the life cycle of an individual (of a specific species), in connection to environmental factors such as temperature and food availability. DEB-based models may be (and have been) implemented into higher-level (e.g., population) models (see general text above).

The available knowledge and data relevant to the risk assessment question

This cannot be answered for the model in general. DEBtox is based on DEB theory, which has a long history in biology.

The outputs required to answer these questions including performance criteria for the regulatory model

This cannot be answered for the model in general. DEBtox can produce various types of useful outputs such as a threshold for effects, EC*x*,*t* for any effect percentage *x* and exposure duration *t*, or the expected growth and reproduction pattern from any (untested) exposure profile, or untested environmental conditions. To expand on the latter: effects of changes in food and temperature have been studies extensively in DEB theory (food level acts by changing *f*, and temperature by affecting all rate constants by the same factor), although much less in conjunction with toxicant effects. Such extrapolations would therefore be accompanied by additional uncertainty (increasing with extrapolation distance).

The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) focusses on extrapolation to other exposure profiles, and proposes a number of criteria for the data and model performance in calibration/validation. The calibrated model needs to be validated for each species-chemical combination, against a series of pulse-exposure treatments, before it can be used for extrapolation to exposure profiles from fate models.

The species to be modelled

DEBtox can be used for many, in principle all, animal species and all chemicals (as well as many non-chemical stresses). However, several species will require model adaptations to accommodate their life cycles. For example, nematodes do not grow according to a von Bertalanffy curve, and species like *C. elegans* stop reproducing quite abruptly. Modifications have been suggested and tested (Jager et al.

2005, Goussen et al. 2015). As another example, holometabolic insects have larval forms that are very different from the adult form, with a radical metamorphosis in between. The best ways to capture such life histories completely in a DEB model is still being investigated in the scientific community (Llandres et al. 2015). For insects, there is currently very little experience in attempting to use DEBtox models to cover the toxic effects. However, the experience with nematodes supports the assumption that the same DEBtox effect modules can be used as long as the basic DEB model is representative.

Requirements for the environmental scenarios to be used in the risk assessment

In terms of environmental scenarios, DEBtox models require specification of the exposure scenario (how the exposure concentration varies over time, e.g., the output from fate models). Temperature and food availability are important factors influencing the energy budget. In toxicity tests, these conditions are generally kept optimal and constant. When fitting on time-varying conditions, or when extrapolating to deviating conditions in a field scenario, the temperature and food profiles over time need to be specified. Furthermore, dealing with different food levels and temperatures places additional requirements on the available data for calibration, or prior knowledge about the species (e.g., how its physiological rate constants depend on temperature, and how external food levels translate into feeding rates). When DEBtox is used to predict effect as a result of fate models (e.g., FOCUS output), the fate models will require environmental scenarios that consider factors such as soil type, rainfall and agronomic practice.

2.3.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

DEB theory is not built on data, but rather on simplifying assumptions about the energetics of organisms. General patterns in life history have been presented as stylised facts to support these assumptions (Sousa et al. 2008). Data is required to calibrate DEBtox, i.e., to fit the model to the data, thereby obtaining values for all of the model parameters and their confidence intervals. Therefore, this item cannot be answered for the model in general. It can only be addressed in the context of a specific risk assessment, a specific data set to be analysed, and when specific protection goals have been defined.

DEBtox can be calibrated using the data from several standard test protocols, as long as observations on growth and/or reproduction are reported over time. However, it can also be calibrated on non-standard data sets, for example, data sets where the exposure concentration was not kept constant over time (e.g., Pieters et al. 2006, Billoir et al. 2011). It should be noted that standard test protocols have not been optimised for the purpose of calibrating mechanistic models. Other test designs (possibly depending on the specific properties of the test chemical and the risk-assessment question) will be far more efficient for the modelling purpose (Barsi et al. 2014). In general, application of DEBtox is best served by determining body size and reproductive output, over time, over a substantial part of the life cycle (starting with juveniles and following them until they approach their final size and have produced sufficient offspring to determine reproduction rates). The 21-day Daphnia test reasonably matches these requirements, when the experimenters follow the OECD recommendation to determine body size at several points in time as well.

Has the quality of the data used been considered and documented?

This item cannot be answered for the model in general. The model itself does not rely on, or contain, data. The only data that are used by the model are data for a specific species or chemical-species combination.

Have all available data been used? If not, is there a justification why this information has not been used?

This item cannot be answered for the model in general. The only data that are used by the model are data for a specific species or chemical-species combination.

2.3.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

This item cannot be answered for the model in general.

Are the modelling endpoints relevant to the specific protection goal?

This item cannot be answered for the model in general. The endpoints of DEBtox are growth and reproduction (often complemented with mortality by adding a GUTS module), or more specifically: the growth and reproduction of an individual organism over time, as a function of an exposure pattern/scenario in the environment. The model is thus relevant for protection goals that involve individual performance. DEB-based models can be linked to, or used in, population models to make projections to population consequences (Jager et al. 2014a).

Is the modelling approach justified?

At this moment, the only TKTD models that have any track record in ecotoxicology for sub-lethal effects in animals are those that are based on DEB theory. This implies that there is no alternative for DEBtox models at the moment, if a dynamic representation of individual growth and reproduction is required. Given the type of data sets that are available for calibration (which generally do not include information on body residues or sub-organismal changes), DEBtox models comprise the maximum level of complexity that can be accommodated.

Is the conceptual model logical?

DEBtox rests on DEB theory, which is logical, consistent and well-tested, and has a long history and considerable distribution in science.

Are the processes included in the model relevant to the addressed issue?

Yes, see section 2.3.1.

Are the links between different processes to the variables logical?

Yes, see section 2.3.1.

Are the temporal and spatial scales relevant in regard to the problem definition?

DEBtox has no spatial scale as such; it focusses on the life history of individuals as a function of food, temperature and stressors. In terms of temporal scale, DEBtox models are generally simulated in continuous time (the model consists of a system of straightforward differential equations). DEB theory considers the entire life cycle, from egg to death, but the standard DEBtox models do not include embryonic development or ageing effects (or interactions between ageing and toxicity), and therefore they are generally limited to the part of the life cycle between birth (start of feeding) and the onset of senescence. Extensions to cover effects on embryos (Barsi et al. 2014) and ageing (Jager et al. 2004,

Alda Álvarez et al. 2005) have been proposed but require further work. For most applications it is assumed that the model parameters remain constant over the time period modelled.

2.3.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

DEBtox has been described in a number of papers in the open literature, but the most detailed (and readable) treatment is in the freely-available e-book (Jager 2019). This e-book also extensively discusses the assumptions underlying the standard DEB animal model and DEBtox models. A list of publications with applications of DEBtox models is maintained at http://www.debtox.info/papers_debtox.html (currently around 90 entries).

Are the most important mathematical equations described?

The mathematical details (including the derivations) are provided in a technical annex to the e-book, as well as in the open literature (Jager and Zimmer 2012).

Is there a description of the variables and parameters including their meaning and unit?

Yes, see above.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

DEBtox is specifically intended to be applied in the absence of detailed information about the toxicity mechanism of the compound and the biology of the species. It only requires data on growth, and preferably reproduction, over a part of the life cycle. This is the type of data that are available from some standardised toxicity tests, although these protocols can relatively simply be adapted to make them far more useful. DEBtox probably constitutes the absolute maximum in complexity that is still acceptable for such types of data sets.

Are references supporting the equations been provided?

Yes, see above.

2.3.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

This item cannot be answered for the model in general, as there are several implementations of DEBtox used by different research groups. The BYOM-Matlab implementation has freely-available and readable Matlab code, and a basic manual.

Is the computer code well readable and is it available?

The BYOM-Matlab implementation, for example, has freely-available and readable Matlab code, and a basic manual.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

The BYOM-Matlab implementation, for example, has undergone critical investigation, the basic structure of the code has been in use for almost 15 years, but no logs have been kept of the checks that have been performed to ensure its correctness.

2.3.3.6 Evaluation of the environmental scenario

Is the scenario representative for the risk assessment under consideration?

This item cannot be answered for the model in general. For DEBtox, the 'scenario' entails the pattern of chemical concentration in the environment over time, possibly extended with temperature and food availability (which are basic forcing variables in DEB). The scenario will thus generally follow directly from the design of the toxicity test that is to be analysed or from the FOCUS profiles that need to be predicted.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

This item cannot be answered for the model in general. See above.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

This item cannot be answered for the model in general. For DEBtox, the 'scenario' entails the pattern of chemical concentration in the environment over time, which is not part of the model itself. DEBtox has no spatial context (it is a model for individuals) and hence there is no 'area' to consider. For the animals, the exposure pathways included are the ones that are available in the toxicity test. Animals are fed in standard chronic toxicity tests, and hence uptake with food will be considered in the model parameterisation and in the predictions.

The toxicity parameters may depend on temperature, food availability, body size, and developmental status. At this moment, there are no clear indications for such effects on toxicity parameters, but the topic has not been studied thoroughly yet. As with all models, extrapolations will become more uncertain with increasing distance between the extrapolation scenario and the situation of the experimental data used for calibration.

Is the level of conservatism placed into the scenarios appropriate?

The scenario is not part of the model, but of a specific application. The model itself is not conservative or non-conservative, it is the scenario (the exposure profile over time) that will determine the level of conservatism (and should lead to selection of an appropriate assessment factor to cover the remaining uncertainties).

2.3.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

This item cannot be answered for the model in general. For DEBtox (like most TK and TKTD models in ecotoxicology, as well as classical dose-response analysis), model parameterisation is the act of calibrating the model to data for a specific case (i.e., toxicity data for growth and reproduction over time, and, if available, body residues as well). In general, it is recommended to estimate the parameters by

maximum likelihood estimation (Jager and Zimmer 2012). Confidence intervals can be calculated with Bayesian (e.g., Billoir et al. 2008a) or likelihood-based (frequentist) methods.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

That has to be established for every application of the model. The data used for parameter estimation will be the results from toxicity tests (or observations on animals over time, when focussing on the basic model parameters). Most standard tests are of limited use for DEBtox parameterisation, although several could easily be modified (Barsi et al. 2014).

Were the estimated parameter values realistic?

This item cannot be answered for the model in general. For the basic metabolic parameters of a species, the parameter values can be compared to those derived in other studies or those from the add-my-pet collection.

Are the data sources sufficiently documented?

This item cannot be answered for the model in general.

2.3.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

DEBtox will be always fitted to toxicity data, which is specific for a chemical and a species. All model parameters obtain their value by fitting the model to a data set, and classical sensitivity and uncertainty analysis thus makes little sense (just as it does not make sense for the log-logistic dose-response curve) (Jager and Ashauer 2018a). All of the relevant information on uncertainty and sensitivity is represented in the parameter estimates and their (joint) confidence intervals. Sensitivity and uncertainty analysis *can* be performed for TKTD models like DEBtox, for any analysis, after the model has been calibrated to a specific data set. However, there are very few examples of such analyses as their usefulness is very limited. Instead of classical uncertainty/sensitivity analysis, it is more useful to focus on a proper statistical treatment for optimisation, and for the subsequent construction of confidence intervals, and the propagation of uncertainties.

If classical uncertainty/sensitivity analysis is required, one would need to consider that sensitivity and uncertainty will depend on the model output (e.g., body size or cumulative reproduction), and change over time and with the exposure treatment. In general, all parameters will be sensitive, and contribute to uncertainty, at some time point under some exposure scenarios (with the possible exception of g, which governs the maximum reserve density).

It is possible to create a basic model for a species, and use that for different chemicals. As an example, the DEBtox Windows software contained defaults for *Daphnia magna* to allow estimation of model parameters in absence of information on body size (Kooijman and Bedaux 1996b). The case studies in the e-book (Jager 2019) provide some insight into the general insensitivity of the parameter g that covers the reserve dynamics. In such cases, some form of sensitivity and uncertainty analysis could add value to a specific model analysis. However, this is technically challenging as there is an optimisation step involved (if we change the value of g, we would need to fit the toxicity data, and make model predictions, again, for every value of g).

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

See above.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

See above. Sensitivity analysis is hardly useful for models that are parameterised by fitting them to data. It is technically feasible to select the 'most-sensitive' parameters after fitting (although this will depend on exposure time and concentration profile), but there is generally little opportunity to refine one of them specifically (individual parameters cannot be measured; they follow from a fit to growth and reproduction data).

Has the uncertainty analysis been adequately documented?

Classical uncertainty analysis is hardly useful for models that are parameterised by fitting them to data. Uncertainties in the model parameters (resulting from the fitting) can be propagated to uncertainties in the model predictions (e.g., Jager and Zimmer 2012).

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

This item cannot be answered.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

Classical uncertainty analysis is hardly useful for models that are parameterised by fitting them to data. We are able to select parameters that contribute most to the uncertainty (although this will depend on exposure time and concentration profile), but there is little opportunity to refine one of them specifically.

Uncertainty is propagated to the model results?

Uncertainty analysis is best covered by propagating the (joint) uncertainty in the parameters estimates to the model predictions. This can be done in every analysis (see e.g., Jager and Zimmer 2012).

It is good to stress that part of the uncertainty in the model predictions will derive from uncertainties in the exposure profile (i.e., the output from the fate models). For a meaningful and consistent propagation of uncertainties, this would have to be done throughout the risk assessment (making the risk assessment probabilistic), and not exclusively for the effect models.

Have confidence intervals been estimated and has this information been used in further model use?

When fitting the model to the data, confidence intervals on the model parameters can be derived in various ways (e.g., asymptotic standard errors, likelihood profiling, Bayesian analysis). The joint confidence intervals on the parameters can be propagated to obtain intervals on model predictions, which can be an EC*x*,*t*, the expected growth and reproduction due to an untested exposure profile, or the intrinsic rate of population increase.

2.3.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

This item cannot be answered for the model in general. The model is completely parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. For each case, specific performance criteria may be defined. EFSA proposes several general criteria for application of TKTD models in ERA for PPPs. However, no strict pass/fail cut offs are provided,

and these criteria are not specifically discussed/demonstrated for DEBtox (more focus is placed on GUTS).

Are the model outputs that are compared relevant in view of the problem definition?

This item cannot be answered for the model in general. TK and TKTD models are generally fitted to data, just like dose-response curves. Therefore, the only criterion for 'validity' can be goodness-of-fit, and whether the model is generally able to provide a good fit to the data (the same is true for the currently-applied dose-response curves). Model validation should be closely linked to the intended purpose of the model, and as DEBtox can be used for different purposes, this item cannot be addressed in general. If DEBtox is to be used to fit experimental data and derive a no-effect concentration or EC*x*,*t*, validation is impossible. If DEBtox is to be used to extrapolate from a data set with constant exposure to a pulsed exposure scenario, validation *is* possible. However, to our knowledge, such validation studies have not yet been published yet, with the possible exception of the prediction of mixture effects from the individual components (Jager et al. 2014b).

It should be noted that such a validation study, in principle, only says something about the validity of the model parameterised for this species and this chemical (and this particular extrapolation). Therefore, the EFSA opinion requires a validation for each case. However, a range of such validations goes a long way towards embodying trust in the general model structure, and provides information on the accuracy/precision of extrapolations which may be used to set reasonable assessment factors. Ultimately, this can lead to modification of the validation requirements for each application. It should be noted that DEBtox rests on the vast amount of work that has been in the framework of DEB theory over the last 40 years, which forms a basis for trusting the general usefulness and applicability of this framework.

DEB models link many aspects of the organism's life cycle. Therefore, the model can be calibrated to one set of data (e.g., growth and reproduction data) and predict other properties (e.g., feeding and respiration rates as function of body size). Examples can be found elsewhere (Jager and Ravagnan 2015, Jager et al. 2017b).

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

See above.

Is the dataset relevant in view of the problem definition?

See above.

Is the fit of model output to the data good enough?

This item cannot be answered for the model in general. The model is parameterised from case-specific data, so independent measurements for model evaluation are also completely case specific. Furthermore, it is unclear what 'good enough' means in this context. A poor fit might be caused by many different factors, and might relate to limitations of the model, as well as to limitations of the experiment (or the link between the two). A poor fit should prompt further investigation (scrutiny of the experimental test and the model calibration, or additional toxicity testing and/or model extension), or an additional assessment factor.

DEBtox is generally used to explain data, and thus fitted. The fact that the model can explain (stressor effects on) very different life-history traits over the life cycle of an organism with one (relatively small) set of parameters is convincing enough for most applications. Therefore, there has never been such a

strict focus on 'validation'. Comparisons between model predictions and independent observations have regularly been presented for DEB models, but they are not referred to as a 'validation' of the model or the theory (more generally as support for the model and its parameterisation).

Has the performance of the model been reported in an objective and reproducible way?

This item cannot be answered for the model in general. EFSA proposes a number of model-performance criteria for validation in GUTS applications, but not specifically for DEBtox.

2.3.3.10 Model Use

Is a user manual available?

Some implementations of DEBtox have a user manual (e.g., BYOM), although more work would be needed if the implementations are to be used by non-specialists.

Have all aspects of the modelling cycle been documented?

For models like DEBtox, there is no simple modelling cycle. Such a cycle is an (overly) simplified representation, that most closely matches the situation where a model is built from scratch by a single research group (or a single person), for a specific purpose (Jager and Ashauer 2018a). DEBtox has not been developed from scratch (it derives from 40 years of work on DEB theory), and has been applied and modified by many different (largely independent) research groups for many different purposes. Therefore, it is impossible to document 'the modelling cycle' of DEBtox in general. For each analysis (for a specific software and species-chemical combination), a sub-cycle might be documented.

Has a summary sheet been provided by the modeller?

These items cannot be answered for the model in general. The previous table is basically the summary sheet for the model in general.

2.3.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

We cannot say whether the modellers who will produce a DEBtox analysis for a particular dossier are available for dialogue. Many scientists are working on the general development and application of DEBtox, and many of them would likely be open to dialogue with risk assessors.

Is a version control system implemented?

DEBtox itself does not have a version control; different versions have been developed and are being maintained by different research groups. Several of the software implementations do have a version control system (such as BYOM). DEBtox is not owned by anyone, nor is there any central coordination on the development, so anyone is free to develop their own version.

2.3.3.12 **Overall Judgement**

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

This item cannot be answered for the model in general. DEBtox can be used for analysis of toxicity data, as a replacement of descriptive and limited dose-response curves. The model has a broad support in the scientific community, and rests on the development of DEB theory over some 40 years. Therefore, the model should, in general, be suitable for regulatory purposes. However, the EFSA opinion points at a lack of published case studies with PPPs and aquatic organisms, and a lack of user-friendly software. Furthermore, there is currently a lack of case studies clarifying the accuracy of extrapolation between exposure scenarios, and lack of guidance on how to derive an EP*x* for a FOCUS profile (following a single daphnid as it grows and reproduces over >1 year would be rather pointless, and not even always worst case). These limitations will hamper routine application of the model in risk assessment, although a good case can still be made for specific cases when substantial data for calibration and validation are available.

DEBtox could be used to extrapolate to other exposure scenarios, e.g., from constant exposure to time-varying exposure. However, all extrapolations are based on assumptions, and at this moment, it is impossible to make claims on the general accuracy and precision of such extrapolations as this aspect has not been seriously tested yet. Since DEBtox is based on a mechanistic representation of TK and TD, extrapolations rest on a solid foundation, which cannot be said for extrapolations using the NOEC or EC*x* from a dose-response curve. Obviously, not all uncertainties can be addressed by the model (and by the available data, see also the list of uncertainties at the end of this evaluation), and an appropriate assessment factor is required.

At this moment, the possibilities for extrapolation between species and between chemicals has attention in the scientific community, but is not at a level where it can be used to reliably predict the model parameters for a new chemical or species for ERA (see Ashauer and Jager 2018). For the basic DEB parameters of species (the parameters that determine the life history in the absence of toxicant stress), the add-my-pet library can be used. This library may be used to derive the basic (non-chemical-specific) DEBtox parameters for untested species. However, extrapolation of the parameters governing the toxic response is more problematic. Insufficient high-quality data sets have been analysed to deduce patterns in these parameters, and there are indications that the physiological mode of action (pMoA) can be different for the same chemical in different species (Alda Álvarez et al. 2006a, Ashauer and Jager 2018). The same can be said for mixture toxicity (Jager et al. 2010, Jager et al. 2014b): DEBtox can be used to analyse the data for mixture effects, and make predictions based on non-interaction of two (or more) compounds, but cannot (yet) be used to predict specific interactions between chemicals (interactions that do not logically follow from the model's structure). It is good to note that the structure of the DEBtox model implies certain inevitable (mainly synergistic) interactions between stressors. The assumption of non-interaction is thus applied at the level of the stressor's target site in the model, and will still lead to interactions at the level of the life-history traits,

On these areas, more structural testing efforts are needed. However, since DEBtox is based on a mechanistic representation of TK and TD, extrapolations rest on a more solid foundation than extrapolations using the NOEC or the EC50 from a dose-response curve. Note that the use of an EC*x* or NOEC is also a model, which shares the uncertainties of a DEBtox analysis, and adds quite a few more (especially those related to time/timing of exposure).

2.3.4 Qualitative Assessment of Uncertainties

The DEBtox models are not by themselves conservative or non-conservative. The level of conservatism depends on how the model is used, which data are used to calibrate it, and which model outputs are used. However, it is possible to identify certain uncertainties that will affect the level of conservatism, which are presented below. An appropriate assessment factor will be needed to ensure that the level of conservatism is acceptable. These factors are quite similar to the ones for methods that are routinely used to deal with sub-lethal effects in ERA at this moment (dose-response analysis, and using EC*x* in conjunction with a time-weighted average or peak concentration). The current methods are also models, and should be evaluated along the same lines as the DEBtox model (Jager and Ashauer 2018a). However, the current methods have a range of additional uncertainties that can lead to more severe under- and overestimation of risk (especially due to ignoring the time aspect of toxicity). DEBtox extracts the maximum amount of information from chronic (sub-lethal) toxicity data, but the information content of such studies is limited. Mechanisms and processes that are not observed in the experimental test would be difficult or even impossible to predict (although the DEB framework allows for meaningful extrapolations to lower food levels and across temperatures).

2.3.4.1 **Potential for Underestimation of Real Risk**

- Short toxicity tests may not reveal additional effect mechanisms that only show up after prolonged exposure to low concentrations (see e.g., Jager et al. 2007). The model cannot be used to predict such effects from short-term data. Extrapolation to longer test durations thus rests on the assumption that the mechanisms observed in the short-term test are also the only ones relevant for the extrapolation scenario.
- Some chemicals can affect the size/quality of offspring (Hammers-Wirtz and Ratte 2000). DEBtox models can be used to analyse such effects, but cannot predict them. Offspring body size is not routinely determined in toxicity tests. Linked to that, several chemicals have been shown to produce more severe effects in subsequent generations (Massarin et al. 2011). Again, these effects cannot usually be predicted from DEBtox from standard test data.
- Toxicity tests are performed under optimal conditions in terms of food, temperature, absence of predators or diseases, absence of other toxicants, etc. As DEBtox focusses on direct effects due to chemical stress only, there is a potential for synergism with other stresses (e.g., if a chemical makes a species more prone to disease). DEBtox can make predictions for effects at limiting food levels and different temperatures (these factors have predictable effects on the life history, within certain ranges), although such extrapolations are accompanied by additional uncertainty. Environmental factors will often automatically lead to (synergistic) interactions with the toxic effect (e.g., Pieters et al. 2006) due to the structure of the model (and which therefore *can* be predicted), but other interactions may occur that cannot be predicted (although this may also include antagonistic ones).
- Some chemicals may specifically exert toxicity during a particular stage of the life cycle (e.g., during embryonic development or metamorphosis). If such a stage was not present in the toxicity test, its specific sensitivity cannot be identified, and the model will not predict it. Some chemicals seem to interact with the ageing process (Jager et al. 2007), causing low-exposure effects late in the life cycle. However, from the perspective of population dynamics, such effects are likely less relevant than effects early in life.
- DEBtox only deals with effects on endpoints that have a clear energetic component: growth, development and reproduction, possibly extended to feeding, respiration and product formation. This implies that DEBtox cannot be used to analyse or predict other types of effects such as changes in behaviour, changes in sex ratio, malformations, etc.

2.3.4.2 Potential for Overestimation of Real Risk

- DEBtox does not include explicit consideration of bioavailability. Bioavailability of many chemicals will generally be lower in a field situation than in a toxicity test. Bioavailability has to be dealt with separately.
- DEBtox does not consider protective behaviour (such as avoidance) or adaptation to the chemical stress (e.g., induction of biotransformation enzymes).

2.3.4.3 Potential for Uncertainty in Either Direction

- DEBtox assumes that it is the internal concentration of the chemical that causes an effect on an energetic process. Therefore, effects of growth are included as changes in the TK rate constants (due to changes in the surface:volume ratio) and growth dilution (decrease of a concentration as the body increases in size). However, toxic effects may relate to some form of damage, for which surface:volume ratio and growth dilution may be irrelevant. As the DEBtox model is always fitted to data sets from toxicity test, it is unclear whether this leads to over- or underestimates of risks in other scenarios.
- DEBtox allows for extrapolations to different food and temperature settings. These extrapolations rest on the assumption that the toxicity parameters do not change with food and temperature. At this moment, there is only limited evidence to support this assumption.
- Calibrating DEBtox for species A and compound B does not imply that the same parameters also hold for other species and other chemicals. The model can be used to predict the results for other species than the test species, by assuming that the chemical-specific parameters are the same in both species. However, this extrapolation is insufficiently tested at the moment.
- The model assumes that effects are reversible when exposure stops, with a rate that is the same as in the accumulation phase. In other words: the parameters established in the calibration are assumed to hold their value for other exposure scenarios. For some chemicals, effects may not be (completely) reversible, making the organism more vulnerable to a subsequent exposure than expected. For others, the organism might have inducible defences that renders them less sensitive to subsequent exposure events. At this moment, there are no DEBtox studies that indicate the importance of these factors. The EFSA opinion recommends validation with pulsed exposure, which should clearly show whether the potential for recovery of the individuals is as predicted from the model.

2.4 IDamP

Evaluation by Tjalling Jager

2.4.1 General Information

2.4.1.1 Background and Concept

IDamP was first published by Thomas Preuss and co-workers in 2009 (Preuss et al. 2009a)[6]. It is an individual-based population model (IBM), not spatially explicit, for *Daphnia magna*. The original paper only deals with the basic life history of *D. magna* in the absence of toxicants; only food stress and crowding effects were considered. Later papers include modules for toxicants, either as static dose-response relationships (Preuss et al. 2010b)[7] or including a dynamic TKTD model for survival from the GUTS framework (Gergs et al. 2013a, Gabsi et al. 2014c, Dohmen et al. 2016). Further studies included more factors such as the effects of temperature on the life history, and the effects of food density on offspring size (Gabsi and Preuss 2014), potential interactions with competitors and predation (Gabsi et al. 2014d), and more detailed predator interactions including effects of kairomones (Gergs et al. 2013a). These extensions are all add-ons to the original model, not entirely new models. The model has been described by the authors as a 'virtual laboratory' for exploring the effects of different stresses (and the interaction of such stresses) without having to resort to complex and difficult-to-interpret experiments. More recently, the model was linked to a complex biogeochemical lake model (StoLaM) to include more ecological realism, such as phytoplankton dynamics, nutrient cycling, etc. This yielded a new combined model DaLaM (see http://gaiac-eco.de/en/modelling/modelling-of-aquatic-mesocosms), which was recently published in the open literature (Strauss et al. 2017).

As with all IBMs, the population part is simply keeping track of all individuals over time. In its basic version (Preuss et al. 2009a), IDamP does not include spatial heterogeneity; the population is living in a homogenous laboratory vessel. The food is treated as inanimate particles, following simple mass balancing: change in algal density is inflow minus what is removed by outflow and feeding. The distinguishing property of IDamP lies mainly in the way it treats the life cycle of the individual. The model rests on an extensive foundation of empirical data and descriptive relationships (i.e., regressions on experimental data). The model was parameterised on a range of data sets for *D. magna*, for life-history traits at different food levels and different stocking levels (crowding). Advantage of such a descriptive approach is that it stays close to the experimental data, without relying on a series of simplifying assumptions about the mechanistic processes underlying the organism's life history. Disadvantage is that it is species specific, requires substantial amounts of experimental data to parameterise for a species, and that extrapolating (far) beyond the environmental conditions of the calibration data could lead to unrealistic results.

Stochasticity is included by allowing individuals in the model to differ in a number of parameter values; they draw random modifying factors from a series of independent normal or uniform distributions at birth. This form of stochasticity is included for the following parameters: maximum filtration rate, growth rate, juvenile development rate, embryonic development rate, brood size, and expected lifetime. Therefore, each model run produces a somewhat different population trajectory. In practice,

[6] Preuss, T. G., M. Hammers-Wirtz, U. Hommen, M. N. Rubach and H. T. Ratte (2009): Development and validation of an individual based *Daphnia magna* population model: the influence of crowding on population dynamics. Ecological Modelling 220(3): 310-329.

[7] Preuss, T. G., M. Hammers-Wirtz and H. T. Ratte (2010): The potential of individual based population models to extrapolate effects measured at standardized test conditions to relevant environmental conditions - an example for 3,4-dichloroaniline on *Daphnia magna*. Journal of Environmental Monitoring 12(11): 2070-2079.

the authors cover this variation by running the model a large number of times (Monte Carlo simulation) and summarising the results as a mean trajectory with minimum-maximum ranges.

To work with chemical stress, additional modules have been added. Initially (Preuss et al. 2010b), this was accomplished by including static dose-response curves for survival and reproduction into IDamP. This implies that, at the onset of exposure, the effect is immediately at the level as that observed at the end of a standard test (2-days for survival, 21-day for reproduction). For the chemical under study (3,4-dichloroaniline) this worked well as reproduction was far more sensitive than survival, growth was unaffected, and the effects on reproduction did not show a clear time dependence (a rather constant percentage effect, relative to the control, which is consistent with a static dose-response curve). The authors recognised that this may be an exceptional case, and expected that for other compounds, more complex effect models at the individual level would likely need to be implemented (i.e., a TKTD model). In later work (Gabsi et al. 2014c), IDamP was coupled to a GUTS model to cover dynamic survival effects of another compound (dispersogen A), but kept the static dose-response approach for reproduction effects. Another study (Gergs et al. 2013a) looked at the toxic effects of nonylphenol and the interactions with predation and predator kairomones, and also applied a GUTS model to deal with mortality due to toxic stress.

2.4.1.2 **Status of the Model**

Development and application of IDamP is currently covered by Bayer (Thomas Preuss) and Gaiac (Tido Strauss), who are also contact persons for this model. The model is not made publicly available, although (part of) the source code was provided with the first version of the model. The model is, however, available from the developers on request, in the form of a standalone software (programmed in Delphi) with a user-friendly interface. A user manual for the software is available as well, and has been made available for our evaluation.

IDamP calculations have been included in a few cases for active-ingredient renewals as well as in dossiers at the national level. In one dossier (bromoxynil-octanoate RAR 2016, Vol. 3-B.9 p. 67-90), the model was applied to extrapolate effects at the individual level from a 21-day constant exposure to a scenario with a single peak of varying duration. This case was evaluated in detail in section 3.4. Even though the model is not particularly suited for such extrapolations, it was attempted to show that the model presents a worst-case prediction. In another case study (for the insecticide pirimicarb, evaluated in section 3.5), IDamP was linked to a lake model to make predictions at the population level for a relevant ecological scenario. An example of how the model could be applied in risk assessment at the population level was published in the open literature by Dohmen et al. (2016), as part of a series of papers resulting from the 2012/2013 'MODELINK' SETAC technical workshops. Currently, nine papers have been published in the open literature that present/apply the model (most of them explicitly mentioning potential for application in ERA). However, these are all resulting from a relatively small user community (focussed around Aachen University).

2.4.2 Model Description

2.4.2.1 Problem Definition

Context in which the model will be used

IDamP is a model for the population dynamics of *Daphnia magna* in a laboratory setting. The model can thus be used to extrapolate effects from standard tests (such as the 21-day reproduction test) to population consequences (at least under conditions similar to those in the laboratory, as used for model calibration). The model has accumulated various extensions over the years which add more ecological realism (e.g., inter-specific competition, predation, see Gergs et al. 2013a, Gabsi et al. 2014d) and has recently been linked to a biogeochemical lake model (StoLaM) for a more realistic environmental scenario (Strauss et al. 2017).

Specification of the question(s) that should be answered with the model

Questions that deal with the extrapolation from individual-level effects to population consequences. The basic model focusses on a single species (*D. magna*) although interactions with other species (predation, competition, algal food) have been included in several studies (and are available in the last version of the software as optional modules). The model has also been applied to provide a (presumably worst-case) extrapolation for individual-level effects from constant to pulsed exposure, although the model is not particularly suitable for such extrapolations (due to the lack of TKTD considerations).

Specification of necessary model outputs and protection goals

Different types of potentially relevant output at the population level may be derived from the model, such as the carrying capacity of the population, population growth rates, time to recovery, duration of adverse effects, risk of extinction, etc. Different options were demonstrated in various papers (Preuss et al. 2010b, Gabsi and Preuss 2014, Dohmen et al. 2016). The model is thus relevant for protection goals that deal with population abundance and biomass due to individual-level toxic effects on mortality, growth and reproduction.

Domain of applicability of the model

The model is parameterised for *D. magna*, and the environmental setting is for a closed laboratory vessel. However, various extensions have been developed to include more ecological realism (incl. the link with the complex lake model StoLaM). The standard model applies a static dose-response curve for the effects on the individuals, which limits its realistic application to fast-acting chemicals. However, the model could be set up provide a worst-case prediction of toxic effects, for slow compounds as well, by excluding recovery of individuals (as done for the pirimicarb case study in section 3.5). However, the model has also been linked to GUTS (Gergs et al. 2013a, Gabsi et al. 2014c, Dohmen et al. 2016), to provide a dynamic calculation of survival effects (but not for sub-lethal effects).

Why is the model being used?

To extrapolate effects on individual life-history traits to population consequences. As a virtual laboratory to explore interactions between various environmental factors (see above) on the population dynamics.

What protection goal is being addressed?

Protection goals aiming to protect aquatic invertebrates (specifically *D. magna*) at the population level. However, the model has also been used for individual performance, extrapolating sub-lethal effects from constant to pulsed exposure (attempting to demonstrate the applicability of a time-weighted average exposure concentration).

What outputs are required?

Different types of potentially-relevant output at the population level may be derived from the model, such as the carrying capacity of the population, population growth rates, time to recovery, duration of adverse effects, risk of extinction, etc. Different options were demonstrated in various papers (Preuss et al. 2010b, Gabsi and Preuss 2014, Dohmen et al. 2016).

How was the species chosen?

D. magna is a standard test species in Tier 1 for ecological risk assessment, and a large body of literature is available on this species.

Which other species/groups are being covered by the chosen one(s)?

The model parameterisation only considers *D. magna*, so no other species or groups are explicitly covered (although *D. magna* can be considered a representative species for zooplankton). However, the model structure could easily be used for other cladocerans, and possibly other algae-feeding zooplankton species, as well. This would require complete re-parameterisation, and hence substantial information on the life history of the species.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Since this evaluation is for the model in general, and not for a specific application in a dossier, only the types of comparisons that have been published so far are discussed. The performance of the model has been evaluated at the individual level and at the population level using laboratory tests. As performance criteria for the individual-level tests, the difference between observed and predicted mean length and cumulative reproduction at the end of the test was used.

For the population analyses, various comparisons were made using laboratory experiments, where a population was followed over time in a closed vessel with regular food supply. In Preuss et al. (2009a), simulated and observed patterns in population abundance (total or divided over size classes) are compared graphically, in absence of toxic stress. In Preuss et al. (2010b), apart from a visual comparison of trajectories, the mean initial population growth rate and final population size were compared between experiment and model prediction (including toxic stress). Furthermore, Preuss et al. (2010b) apply an 'area comparison', to see which fraction of the data points is within the minimum-maximum bounds of the model predictions. The combination of IDamP with the lake model StoLam has been compared to data from mesocosm studies (Strauss et al. 2017).

2.4.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

The model is based on extensive experience with *D. magna*, and data sets on life-history traits in various settings, originating from the University of Aachen. The model was tested using individual-level and population-level laboratory experiments.

Assessment of the quality of the data

The quality of the data is not addressed in the publications of the model. All data sets originate from the same laboratory, which helps ensures consistency, but may not represent the variability between laboratories. References are provided for the data sets used; part of the data used originate from PhD theses that may be difficult to obtain.

2.4.2.3 Conceptual Model

Description of the model concepts including a diagram

A flow chart of the individual's life-history scheduling was provided in Section 2.4.1.1. The model follows a rather common strategy for IBMs: use extensive empirical information on the life history of individuals to construct descriptive relationships, and subsequently follow all individuals in a simulated environmental setting.

Identify the main components and processes in the system

The life-history of individuals is captured by a large set of empirical relationships (i.e., regressions on data), including some semi-mechanistic relations. To exemplify the latter: body length is a state variable, which was used as explanatory variable in a regression of feeding rate, which in turn is used as a factor to calculate brood size.

How the effects of the chemicals are modelled

Static dose-response relationship, or (for survival only) link to a TKTD model (GUTS). How GUTS is exactly implemented seems to differ between different studies. For example, in Gergs et al. (2013a), only the IT model is implemented, with TK following a somewhat odd dependence on body size (and implicitly assuming that damage repair is fast). In contrast, Gabsi et al. (2014c) apply both SD and IT, with size-independent TK/damage dynamics.

How the components and processes are linked

See diagram in Section 2.4.1.1.

2.4.2.4 Formal Model

Identification of the model variables

For the first version of the model the state variables were summarised by the authors in their Table 1 (Preuss et al. 2009a).

Identification of the model parameters

For the first version of the model, the model parameters were summarised by the authors in their Table 3 (Preuss et al. 2009a).

Description of the most important model equations or algorithms

For the first version of the model, the model equations for the individual were summarised by the authors in their Table 2 (Preuss et al. 2009a).

2.4.2.5 Computer Model

Description of the model implementation

The software of the model is not publicly available. However, the software (as standalone executable, programmed in Delphi) and a manual are available from the developers on request.

Checking the computer model for errors, bugs and inconsistencies in the code

According to the developers, the model implementation has been intensively tested, as explained in a summary sheet in an example dossier that was made available to us for evaluation (modelling report for bromoxynil-octanoate). For example, the model was re-implemented by different research groups in different software (C++ and NetLogo), and a series of extreme cases were simulated to see if the model predictions make sense.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

See above.

2.4.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context **in which the model is run**

Different scenarios can be run for the model. In the basic model, the scenario will be a population of *D. magna* living in a laboratory vessel with regular supply of algae. Several scenarios for (time-variable) toxicant exposure and (time-variable) environmental conditions (e.g., food supply and crowding) can be simulated. The model has been extended with various options to include more ecological realism, and has even been linked to a complex lake model (StoLaM).

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

The abiotic parameters (temperature, vessel volume, flow rate) should be set to appropriate values (e.g., to mimic the circumstances under which the validation data will be generated). Biotic parameters are set to represent the life history of *D. magna* (see previous items). Modules have been proposed to include other abiotic processes such as predation and inter-specific competition, which would need to be parameterised as well.

The standard model does not involve agro-environmental parameters, apart from the concentration profile to which the animals will be exposed (if the model is forced by a FOCUS exposure profile, that profile will follow from a scenario with environmental and application conditions). When IDamP is linked to the lake model StoLaM, the agro-environmental setting will become part of the model itself and need to be parameterised.

2.4.2.7 Parameter Estimation

Description of the model parameter estimation

The parameter estimation involved regressions on a series of experimental data sets at the individual level; the parameter values are thus regression coefficients. This also holds for the parameters governing intraspecific competition (crowding) and toxic effects.

Parameters estimated from the literature — what are the sources and why are these appropriate?

Most of the data used were taken from experiments conducted at the University of Aachen (references are provided in the publications of the model).

Parameters obtained from calibration — how and why this was done?

All parameters were estimated at the individual level (by regressions on experimental data); there was no calibration of the population model to any data at the population level.

2.4.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

No classical sensitivity analysis was performed, as this was deemed unnecessary by the authors (as all of the parameters were established by regressions on the experimental data at the individual level). This is a sensible argumentation; see also discussion in Jager and Ashauer (2018a).

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

No classical uncertainty analysis was performed. However, variability between individuals is included stochastically, and is propagated to the model output. The initial values for the state variables are randomly selected from probability distributions at birth, for each individual (and kept fixed throughout the individuals' lifespan). These distributions were based on the data sets at the individual level as used for model parameterisation. However, it is not well documented how these distributions were derived.

2.4.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

The basic model (without toxicity modules) has been compared to independent data in Preuss et al. (2009a) by comparing model predictions to measured data, both at the individual level (for a food situation not used for parameterisation) and the population level (two different experiments: flow-through and semi-batch feeding). In Preuss et al. (2010b), several population experiments with a chemical stress (3,4-dichloroaniline) were used for comparison. In Gabsi et al. (2014c), population experiments with dispersogen A were used. In Gergs et al. (2013a), model predictions were compared to population experiments including predators (backswimmers) and pulses of nonylphenol. For the combination of IDamP with the lake model (DaLaM), model predictions were compared to data from outdoor mesocosm studies (Strauss et al. 2017), without toxicants, and one case study with a toxicant (more detail in the evaluation of the case study with pirimicarb in section 3.5).

Demonstration that the model output provides an adequate match to data patterns

The model provides a very reasonable correspondence with the data (probably as good as can be expected from a population model for *Daphnia*). Also, the ability of the combined model (DaLaM) to predict *Daphnia* density in outdoor mesocosm was generally good (several studies performed less well, which could be explained by factors not considered in the model such as predation by backswimmers). However, the data sets used for 'validation' clearly cannot cover all relevant aspects of *Daphnia* population dynamics. For a specific application of the model, it should be considered whether the relevant aspects were properly covered in validation.

2.4.2.10 Model Use

Explanation of how the model conforms to the requirements set in the problem definition

This item cannot be addressed in general. The model follows a simple principle whereby all individuals are followed (IBM), and the life history of each individual is based on experimental data for the species.

Description how the model works (user manual).

The model has been described in a range of papers. A manual for the standalone software is available.

Description of the pesticide parameters values used in the model

This item cannot be addressed in general. Pesticide effects are included as static dose-response curves for the relevant endpoints. For effects on survival, GUTS can be used (it is currently implemented as an option).

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed in general.

2.4.3 Model Evaluation

2.4.3.1 Problem Definition

The regulatory context in which the model is run

IDamP is a model for the population dynamics of *Daphnia magna* in a laboratory setting. The model can thus be used to extrapolate effects from standard tests (such as the 21-day reproduction test) to population consequences (in principle for *D. magna* in isolation). The model has accumulated various extensions over the years which add more ecological realism (e.g., competition, predation) and has recently been linked to a biogeochemical lake model for a more realistic environmental scenario (Strauss et al. 2017).

The question that has to be answered with the model

The basic IDamP (Preuss et al. 2009a) is a population model for a single species (*D. magna*), in a closed (laboratory-scale) environment, in interaction with several environmental conditions. Initially, the model was restricted to a small number of environmental factors: food (algae) density and conspecifics (crowding, intra-specific competition for food). Later studies have added more ecological factors such as competitors, predation pressure, temperature and toxicant stress. The algae are not treated as a population (they are considered 'food particles'), and neither are the predators, which have been included as a mortality rate (Gabsi et al. 2014d), or in a more detailed mechanistic manner focussing on the feeding behaviour of an individual predator (Gergs et al. 2013a). Inter-specific competition has been treated as an additional dynamic population of daphnids with slightly different properties. In the current version of the software, these additional modules can be switched on or off. More recently, IDamP was combined with a biogeochemical lake model to include more ecological realism in the environmental scenario (Strauss et al. 2017), including treatment of the phytoplankton as a dynamic population.

The model is not developed to answer one specific risk assessment question. Different types of potentially relevant output at the population level may be derived from the model, such as the carrying capacity of the population, population growth rates, time to recovery, duration of adverse effects, risk of extinction, etc. Different options were demonstrated in various papers (Preuss et al. 2010b, Gabsi and Preuss 2014, Dohmen et al. 2016). The model has also been used in the interpretation of effects at the individual level (Agatz et al. 2013). In one dossier (bromoxynil-octanoate, see evaluation in section 3.4), IDamP was used at the individual level, in an attempt to show that the chemical had negligible effects on growth and reproduction under realistic exposure conditions (i.e., extrapolating from constant to pulsed exposure). Negligible effects at the individual level generally imply negligible effects at the population level as well. In another dossier (modelling report for pirimicarb, see evaluation in section 3.5), IDamP was linked to a lake model and used to predict population consequences in a mesocosm setting. A published example of how the model can be applied in risk assessment for impacts at the population level was presented in Dohmen et al. (2016).

The available knowledge and data relevant to the risk assessment question

This item cannot be answered for the model in general.

The outputs required to answer these questions including performance criteria for the regulatory model

This item cannot be answered for the model in general. See above for the potential outputs from the model. Outputs like population growth rates, time to recovery, duration of adverse effects, and extinction risk may certainly be relevant for regulatory purposes.

The species to be modelled

IDamP is specific for *D. magna*, although the same model structure can be used to model other zooplankton species if sufficient information on the life history (and the influence of food, crowding, and other relevant factors) is available.

Requirements for the environmental scenarios to be used in the risk assessment

The environmental scenario used in the model comprises the food situation (how algae are introduced into the system, e.g., flow through or semi batch) and the water volume the Daphnia live in (which is important to include effects of crowding). The model applies an artificial setting of a Daphnia population growing in an enclosed (laboratory) environment with addition of algae (which themselves are not treated as a population). This aids comparison of the model predictions to results from laboratory experiments (allowing straightforward output corroboration), and the identification of general mechanisms in the action of stressors on population dynamics, at the expense of limited ecological realism. To include more field-relevance, IDamP has been linked to the lake model StoLaM.

2.4.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

For the supporting data, three categories of data need to be distinguished. There are data used to derive the relationships that govern the basic life history of the Daphnia (model parameterisation), such as the relationship between body length and filtration rate. These data can be considered part of the model as they will generally remain the same for a range of applications. Additional data would be needed to include the effects of a toxicant on the life-history traits. For these data, their quality can only be addressed for a specific case. And finally, data were used for 'validation' (comparison to independent data, both at the individual level and the population level). Overall, the data are representative and quite extensive, both for the parameterisation (at the individual level) and for testing of the model (individual and population level). However, it should be clear that not all potentially relevant ecological settings have been (and can be) covered by the calibration and validation data sets.

Has the quality of the data used been considered and documented?

The papers presenting the model (Preuss et al. 2009a) and the extension to toxicants (Preuss et al. 2010b) do not discuss the quality of the data, and do not document it. References are provided for the data used, although many of these references are PhD theses that may be difficult to obtain (and are in German). The data are clearly relevant for the model development and testing. In the case study by Gabsi et al. (2014c), an extensive TRACE document was added as supporting information, which also includes a treatment of the data used for parameterisation and testing of the model. A discussion on the quality of the data, and a justification of why these particular data were used, is lacking.

Have all available data been used? If not, is there a justification why this information has not been used?

Clearly, not all available data have been used, as countless more experiments have been performed with *D. magna*. There is no justification why these particular data sets were selected, but they all seem

to originate from the University of Aachen and are clearly relevant. This helps ensure consistency between the various data sets, but raises the question how representative these results are for other labs and other clones of Daphnia.

2.4.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

This item cannot be answered for the model in general.

Are the modelling endpoints relevant to the specific protection goal?

This item cannot be answered for the model in general. The model is capable of producing endpoints that would be relevant to specific protection goals (e.g., population growth rate, time to recovery, duration of adverse effects, extinction risk).

Is the modelling approach justified?

The model follows a rather common strategy for IBMs: use empirical information on the life history of individuals to construct descriptive relationships, and subsequently follow all individuals in the simulated environmental setting. The logic of the conceptual model is thus to focus closely on experimental data at the individual level. The life cycle of the individual, and the effects of toxicants on the traits, are captured by regressions on experimental data. This is done in a very detailed way, and the various traits are linked semi-mechanistically. For example, the individuals grow continuously over time, and filtration rate is linked to the individual's size. Furthermore, juvenile development rate, body growth and reproduction rate depend on food availability. This leads to an accurate description of growth and reproduction over time at different food levels. The focus on empirical data is logical and justified as long as the model application does not require extensive extrapolation to conditions outside of the scope of the data used for parameterisation. There is no explicit consideration of mass and energy conservation; some extrapolations could thus lead to unrealistic behaviour of the model for individuals.

Toxicant effects are included as static dose-response curves, apart from effects on mortality where an optional GUTS module was proposed more recently. Lack of TKTD considerations would make the model most suitable for chemicals with rapid kinetics/dynamics, such that the effect percentage closely follows the external exposure (on constant exposure, this implies a constant EC*x*). However, a worst-case model prediction can be produced by disabling the possibility for recovery at the individual level (as done for the pirimicarb dossier, see section 3.5). Using a static dose-response curve for reproduction is defensible if there is no clear time dependency of the effects and if there are no effects on body size (Preuss et al. 2010b). If there are effects on body size, there will be effects on reproduction as well (as body size determines feeding rates, which also affect reproduction). In such a situation, it would not be consistent to apply independent dose-response curves on growth and reproduction. It is unclear how such effects should be included into the model.

The different stressors (food limitation, crowding, temperature, toxicants, etc.) are assumed to act independently on the individual, which might not be (entirely) correct, especially as the individual is represented by a series of largely empirical relationships. Each stressor is quantified on experimental data dealing with this stressor in isolation; no data with multiple stressors acting simultaneously are used for parameterisation (although some interactions are included in the validation, such as those between chemical stress and crowding). Most of the interactions emerging from the model occur at the population level. The validation studies show that, in the cases treated, the omission of potential interactions between processes at the individual level (treating all processes as independent) does not lead

to unrealistic population predictions. At least, this holds for the ecological settings and compounds applied in the validation studies, which cannot cover all situations potentially occurring in the field.

Is the conceptual model logical?

See above.

Are the processes included in the model relevant to the addressed issue?

See above.

Are the links between different processes to the variables logical?

See above.

Are the temporal and spatial scales relevant in regard to the problem definition?

The spatial scale of the model represents a laboratory setting: a homogeneous vessel of up to a few litres. However, combining IDamP with a lake model (StoLaM) allowed simulation of mesocosm conditions (Strauss et al. 2017). Regarding the temporal scale, feeding is considered in steps of one hour, and the other processes in steps of one day. A time step of one day is rather coarse, and should be carefully considered when dealing with rapidly-changing conditions. The model can be run indefinitely, but the simulations span several months in the published case studies. Whether this is (sufficiently) relevant for a risk assessment depends on the specific protection goals and the regulatory question at hand. The focus on a simple environmental setting has the advantage that the model remains transparent, that results are easy to interpret, and that the observed interactions can be understood from the underlying processes. Disadvantage is that many potentially important ecological factors are not considered in the model (e.g., spatial heterogeneity, food-web interactions, nutrient recycling, migration, disease). Linking IDamP to more detailed models for the environment and phytoplankton can increase ecological realism (Strauss et al. 2017), also by including a more relevant spatial scale, at the cost of added complexity and data needs.

2.4.3.4 **Formal Model**

Are the most important model assumptions justified by the modeller?

The presentation of the model in the open literature (Preuss et al. 2009a) includes an extensive model description, including the equations for the behaviour of the individuals and a list of parameters with their values and units. An overview of the model following the ODD protocol is included in Preuss et al. (2009a) and Gergs et al. (2013a). For one paper, with application to a chemical, a TRACE document was prepared (Gabsi et al. 2014c). The equations for the individual behaviour are explained and discussed, and supported with references where needed. However, the equations are presented in a rather awkward manner in a number of tables (probably to stick to the journal format), which makes them difficult to follow. There is no justification for the level of complexity of the model.

The model is not based on mechanistic assumptions regarding the life history of the species, but rather on a description of the traits (as function of body length, food availability, crowding, etc.) with empirical regressions. The most important assumption is thus that the patterns observed in the calibration data also hold for the scenarios used for the population predictions. This is reasonable as long as the simulated scenarios are not too different from the conditions in the calibration experiments. Implicit assumption is that the various environmental factors act independently at the individual level. This is

clearly a source of uncertainty in the model, although the (still limited) validation studies do not point at gross violations of this assumption.

Are the most important mathematical equations described?

Yes, see above.

Is there a description of the variables and parameters including their meaning and unit?

Yes, see above.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

There is no justification for the level of complexity of the model.

Are references supporting the equations been provided?

Yes, see above.

2.4.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

The software of the model is not publicly available. However, the software and the manual are available from the developers on request. The code of the first version of the model (Preuss et al. 2009a) is quite readable and is available for download as supplementary material to the original paper. The software is standalone and has a user-friendly interface. Note that there are multiple versions of the model, as the model has been extended in subsequent papers. These extensions are all add-ons to the original base model (they can be switched on or off in the last version of the software). For these extensions, no code is publicly available.

Is the computer code well readable and is it available?

See above.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

According to the developers, the model implementation has been intensively tested (explained in a summary sheet in the modelling report for bromoxynil-octanoate). For example, the model was re-implemented by different research groups in different software (C++ and NetLogo), and a series of extreme cases were simulated to see if the model predictions make sense.

2.4.3.6 The Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

That depends on the 'risk assessment under consideration'; the model allows for simulating many different scenarios. The environmental setting used for model development is rather artificial, as the basic model considers a population of *D. magna* living in isolation in a laboratory vessel with a regular supply of algae. However, different scenarios can be simulated with the model, including more ecological realism, which has been demonstrated in a range of publications. Furthermore, the model has been

linked to a lake model to provide a more realistic environmental setting (at the expense of more parameters and complexity of interpretation). However, it is unclear how the lake model relates (or can be related) to the FOCUS scenarios that are generally used for risk assessment.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

See above.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

See above. There is no area under consideration. For the animals, the exposure pathways included are the ones that are available in the toxicity test, as the dose-response curve from the test is used as is.

Is the level of conservatism placed into the scenarios appropriate?

The model itself is not conservative or non-conservative, it is the choice of scenario, and the type of application (see item 1), that will eventually determine the level of conservatism. See also the discussion of uncertainties at the end of this evaluation. A suitable assessment factor would need to be selected to ensure an overall acceptable level of conservatism.

2.4.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

The parameter estimation involved regressions on a series of experimental data sets; the parameter values are thus regression coefficients. These regressions have been documented in the various papers (as listed above), and the data were well suited for this purpose. However, it is not entirely clear how the parameter values of the basic model and their ranges (Table 2 and 3 in Preuss et al. 2009a) relate to the regressions (Fig. 3 in the same paper). For example, Table 3 lists a length at birth of 0.75 mm, while in Fig. 3C the growth curves start at >1 mm. Clearly, the procedure was not sufficiently documented to reconstruct the parameterisation.

Data sources are referenced; most of the data used were taken from experiments conducted at the University of Aachen. Most data are taken from PhD theses in German, and therefore not easily accessible. Whether the parameterisation is sufficient or not depends on what the model is being used for. Clearly, extrapolations beyond the range of the experimental conditions in the parameterisation data will be accompanied by uncertainty (which is true for all methods relying heavily on data, including the methods that are currently routinely used for risk assessment).

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Yes, see above.

Were the estimated parameter values realistic?

Yes, see above.

Are the data sources sufficiently documented?

Yes, see above.

2.4.3.8 **Sensitivity and Uncertainty Analysis**

Has the sensitivity analysis been adequately documented?

No sensitivity analysis has been performed, as this was deemed unnecessary by the authors (as all of the parameters were established by regressions on the experimental data at the individual level). There is thus no calibration on population-level data, or default parameter guesses, for which a sensitivity analysis would have been particularly useful. Therefore, the lack of a classical sensitivity analysis is defensible.

In various papers (as listed above), the effect of ecological factors on the population dynamics, alone and in interaction with each other and with other factors such as toxic stress, have been simulated. This can be viewed as a form of structural sensitivity or uncertainty analysis (investigating how model output depends on the addition of several processes and hence the structure of the model).

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

No sensitivity analysis performed, see above.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

No sensitivity analysis performed, see above.

Has the uncertainty analysis been adequately documented?

Parameter uncertainty is, strictly speaking, not quantified or propagated. However, variability between individuals is included stochastically, and propagated to the model output. At birth, each individual receives a set of independent random factors modifying maximum filtration rate, growth rate, juvenile development rate, embryonic development rate, brood size, and expected lifetime. No uncertainty or variability is assumed in the effects of crowding. The distribution used for each factor was based on the observed variation in the available experiment data. However, it is not clear from the publications (nor from the TRACE document) how this was done exactly. Deriving meaningful distributions for inter-individual variation from regressions (apparently using mean values) is not trivial. Furthermore, the data on the different factors are not independent in the observations, e.g., part of the variation in the observed body size comes from the variation in the filtration rate, and hence the ingestion rate. Assigning the observed variation in body size to the growth process, and using variation on filtration as well, would exaggerate the total variation in the model predictions (as part of the variation is included twice). It is unclear whether these various sources of variation are separated, but the different distributions presented are applied in the model as if they are independent.

This is a procedure that is closely related to uncertainty propagation, but slightly different: it is not the uncertainty in the model parameters that is quantified, but the various regressions at the individual level are used to construct distributions for inter-individual variation. This variability is propagated to the population development over time. As the parameters for each individual are selected randomly, each model run yields a somewhat different population response. A range of model runs (usually 100-1000) is therefore performed, which is summarised into a mean trajectory and minimum-maximum intervals on model outputs. In the validation tests, the predicted variation was reasonably consistent with the observed variation in the experimental data at the population level.

This latter part of the approach is well documented. However, the first part (derivation of the distributions for the individuals) is insufficiently documented, which makes it unclear what the confidence intervals on the model output actually represent. At the level of the individual life-history traits (e.g., Fig.

7 in Preuss et al. 2009a), the predicted variability seems exaggerated. Unfortunately, there is no reality check for the predicted variability; it seems that the predicted life-history traits (with bounds) are only compared to mean observed responses from a group of individuals (in the bromoxynil RAR, there is one plot that may show individual observations on reproduction – Fig. 7-1 – and the inter-individual variability was clearly overestimated by the model).

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

This item cannot be answered.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

There has been no identification of the parameters contributing most to the overall variation, although it should be possible to add such a calculation if needed. This would be specifically useful when the level of uncertainty in the model analysis would need to be refined (i.e., when the confidence bands are too wide).

Uncertainty is propagated to the model results?

To some extent, see above.

Have confidence intervals been estimated and has this information been used in further model use?

To some extent, see above.

2.4.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

No performance criteria predefined.

Are the model outputs that are compared relevant in view of the problem definition?

There is no general 'problem definition'; the model can be used for various purposes. The basic model has been compared to independent data in Preuss et al. (2009a) by relating model predictions to measured data, both at the individual level (for a food situation not used for parameterisation) and the population level (two different experiments: flow-through and semi-batch feeding). In Preuss et al. (2010b), several population experiments with a chemical stress (3,4-dichloroaniline) were used for comparison. In Gabsi et al. (2014c), population experiments with dispersogen A were used. In Gergs et al. (2013a), model predictions were compared to population experiments including predators (back-swimmers) and pulses of nonylphenol. The data sets were taken from the literature, and references have been provided. In one case, the experiments were a dedicated part of the study (Gergs et al. 2013a). The combination of IDamP with the lake model was compared to data for Daphnia abundance in outdoor mesocosms (Strauss et al. 2017), setting up the lake model to mimic mesocosm conditions.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

No discussion on data quality is included, but references are provided. The data seem relevant for the model setting.

Is the dataset relevant in view of the problem definition?

The data sets are relevant for the model purpose.

Is the fit of model output to the data good enough?

The model provides a very reasonable correspondence with the data (probably as good as can be expected from a population model for Daphnia). Also, the predicted variation was reasonably consistent (generally somewhat exaggerated) with the observed variation in the experimental data at the population level (Preuss et al. 2009a). Furthermore, effects of toxicants on the population dynamics (abundance and size structure) were also reasonably well captured (see references in the previous items). There is no way to judge whether the correspondence is 'good enough' as that would require a specific purpose and a set of detailed performance criteria.

Has the performance of the model been reported in an objective and reproducible way?

As performance indicators for the individual-level tests, the difference between observed and predicted mean length and cumulative reproduction at the end of the test was used. For the population analyses, the mean initial population growth rate and final population size were compared between experiment and model prediction. Furthermore, an 'area comparison' has been suggested (Preuss et al. 2010b), to see which fraction of the data points is within the minimum-maximum bounds of the model predictions.

2.4.3.10 Model Use

Is a user manual available?

The model and user manual are not publicly available, but can be requested from the contact persons. The manual was made available for our evaluation. The model is a standalone Windows executable with a user-friendly interface (as judged from the screen shots in the manual). The manual seems to adequately describe the basic workings of the software.

Have all aspects of the modelling cycle been documented?

The original presentation of the IDamP (Preuss et al. 2009a) treats most aspects of the modelling cycle for the basic model without toxicant stress. There is no explicit discussion of all the elements in the modelling cycle, and there is no single modelling cycle: there have been nine papers on IDamP, each dealing with a different version of the model, and each addressing different questions. Many of the elements in the modelling cycle are discussed in the published papers, but not in a formal manner. There is a presentation of the model following the headings of the ODD protocol for IBMs in (Preuss et al. 2009a) and Gergs et al. (2013a). A website is available with a cursory overview of the model and information on its history and applications: http://www.bio5.rwth-aachen.de/index.php/forschung/modellierung-und-simulations/27-idamp-model. In the case study by Gabsi et al. (2014c), an extensive TRACE document was added as supporting information.

Has a summary sheet been provided by the modeller?

For one dossier available to us (bromoxynil-octanoate, modelling report), a summary sheet was provided by the modellers according to the template in the EFSA Sci. Op. on GMP (Table 1 in EFSA PPR 2014b).

2.4.3.11 **Suitability of the Model for Regulatory Purposes**

Is there a possibility for dialogue between the modeller and the risk assessor?

Development and application of IDamP is currently covered by Bayer (Thomas Preuss) and Gaiac (Tido Strauss), who are the contact persons for the model. A publicly-available example of how IDamP could be used for risk assessment of plant protection products was presented by Dohmen et al. (2016).

Is a version control system implemented?

The developers keep a version log for the model, and all versions for all publications have been stored.

2.4.3.12 **Overall Judgement**

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

It is difficult to answer this question for the model in general, as it is unknown how the model will be used, for which regulatory purposes, what scenarios will be simulated, and what chemical-specific data are available for parameterisation and validation. In general, the model is suitable to address regulatory questions that concern Daphnia population dynamics. As such, it provides more information on the toxicant effects than the EC*x* or NOEC from a standard toxicity test. In the two case studies (for specific dossiers in sections 3.4 and 3.5), a more detailed discussion on suitability was possible. In general, the model is well documented, although several aspects (especially the parameterisation and the quantification of individual-level variability) are insufficiently described to allow for a reconstruction.

It is important to note that the basic model only considers a single species, living in isolation in a laboratory environment. The model can be used to extrapolate from individual-level toxicity tests to simple population settings. It could also be used to extrapolate individual-level toxicity to different exposure scenarios, e.g., constant to pulsed exposure. However, TKTD models would be more appropriate tools for that application. IDamP applies static dose-response curves, which make the model most appropriate for chemicals where the level of effect closely follows the external exposure concentration. However, the model can be extended with a TKTD model to deal with effects on survival (GUTS), or turned into a worst case by excluding all possibility for individual recovery (as done for the pirimicarb dossier).

The model has been extended with several environmental and ecological factors, and can be seen as a virtual laboratory for exploring interactions between various factors and the population dynamics. Clearly, in a field situation, there will be far more factors than can reasonable be included into such a model (or any model, for that matter). However, the model includes more ecological realism than the use of an EC*x* or NOEC, and provides more insight into the underlying causes of effect patterns than a mesocosm or field study.

IDamP focusses on a rather simple environmental setting, with the advantage that the model results can be readily interpreted and properly compared to observed population dynamics in laboratory setups. The disadvantage is a limited ecological relevance. However, IDamP can be combined with more complex biogeochemical models to provide a closer link to field conditions (Strauss et al. 2017).

The model has already been submitted for regulatory purposes although it is unclear what specific questions have been addressed (in one case, it was used to extrapolate individual-level toxicity from constant to pulsed exposure). It has a fair track record in the scientific literature (nine papers). However, the user community is rather small, and related to a single research group (Aachen University).

2.4.4 Qualitative Assessment of Uncertainties

The IDamP model is not by itself conservative or non-conservative. The level of conservatism depends on how the model is used, which optional processes are included, which chemical-specific data are used to parameterise and test it, which model outputs are used, and what scenario is run. We can, however, identify certain areas that will affect the level of conservatism, which are presented below. This level of conservatism has to be considered in an overall risk assessment.

2.4.4.1 Potential for Underestimation of Real Risk

- The behaviour of the individual is based on regressions to experimental data. Therefore, the effects included are limited to what has been observed in (rather short-term) experimental data (under laboratory conditions), looking at one factor at a time (either crowding, or food limitation, or toxicant stress, etc.). Longer exposure, or presence of multiple stressors, may reveal stronger effects. However, this can be viewed as a limitation of the parameterisation data sets, rather than a limitation of the model concept (additional experiments can be designed to investigate effects of prolonged exposure or simultaneous exposure to multiple stressors).
- The basic model considers a population in a closed environment with a regular and constant supply of algae. As a consequence, the population will grow to a carrying capacity which is determined by the food level (and the crowding effects). In such a setting, various types of toxic effects on the individual will hardly affect the equilibrium population density (see e.g., Martin et al. 2014). For example, even very strong toxic effects on reproduction may disappear in the equilibrium situation as there is very little reproduction anyway, close to the carrying capacity. However, under different conditions, effects will become visible, e.g., when the population is kept in check by predators or when the population is responding to an algal bloom or recovering from a toxicant pulse. There is thus potential for underestimation of risks, but this is mainly a matter of selecting appropriate scenarios (with a range of ecological conditions that can reveal different aspects of toxicity), using appropriate model outputs (e.g., not only looking at population abundance at carrying capacity), and a careful definition of the specific protection goal.

2.4.4.2 Potential for Overestimation of Real Risk

- IDamP does not include explicit consideration of bioavailability. Bioavailability of many chemicals may be lower in a field situation than in a toxicity test. Bioavailability has to be dealt with separately. Clearly, this is not an issue that is specific for this model; it holds for almost all models, and also for the currently applied methods (e.g., use of Tier 1 tests).
- The model considers a closed system, so there is no potential for recovery through migration of individuals from other areas.
- The original version did not include TK and TD considerations for the individual, but used static dose-response curves. This is representative for fast-acting compounds but would lead to bias for slower-acting compounds. The model has been extended with a GUTS module to capture dynamic effects on survival in several papers, which increases the realism. However, the model does not include TKTD considerations for sub-lethal effects; static dose-response curves are the only option. This will likely represent a worst-case assumption in many situations, as effects will start immediately with the start of exposure. However, it also implies instantaneous recovery of individuals. In some cases, likely depending on the exposure pattern, the use of a static dose-response could also lead to an underprediction of risk. If rapid recovery at the individual level cannot be demonstrated for a compound, recovery could be turned off (for individuals) in the model to produce worst-case estimates without the need to resort to a TKTD model for sub-lethal effects as well (this was actually done in the dossier for pirimicarb).

2.4.4.3 **Potential for Uncertainty in Either Direction**

- The model deals with a population of *D. magna*, living isolated in a closed, homogeneous environment, feeding on one algal species. The real world is obviously more complex, including more species, more stressors, and more interactions. Some simulation work with predation and competition was performed (Gabsi and Preuss 2014, Gabsi et al. 2014d), as well as linkage to more complex models for the environment (Strauss et al. 2017).
- The model parameterisation relies heavily on the available experimental data (there is little theoretical basis). If the population extrapolation deals with different conditions (or in a different combination of conditions) than those in the experimental tests, this can cause bias in either direction. The eco(toxico)logical setting for the model predictions thus have to be considered in light of the conditions for the experimental studies used in model parameterisation (extrapolation beyond the tested conditions will come with substantial uncertainty).
- The various stressors are assumed to act independently on the individual (although interactions will emerge at the population level). In practice, synergy or antagonism may occur between stressors at the individual level that would not be predicted in the absence of direct experimental evidence. However, in the (still limited) validation experiments, with and without toxicants, there were no indications for such unidentified interactions: crowding, food stress and chemical stress have acted simultaneously on the experimental animals, and did not produce strong deviations from the model predictions assuming independent effects. Additional experiments with combinations of stressors at the individual level could provide more support for the model's ability to accurately represent the response to such conditions (or suggest model improvements).
- Parameterisation focussed on a single clone of *D. magna*, and the data originate from the same laboratory. Other clones (in other labs) may respond somewhat differently. Furthermore, the model only considers *D. magna*; other species may respond very differently to toxic stress.

2.5 IBM *Chaoborus* Population Model

Evaluation by Jeremias Becker

2.5.1 General Information

2.5.1.1 Background and Concept

The IBM *Chaoborus* population model is an individual-based model that simulates the life cycle of the phantom midge *Chaoborus crystallinus*. The model has been published by Strauss et al. (2016)[8] to investigate effects of cannibalism on the population dynamics and for the extrapolation of toxic effects from individuals in the laboratory to populations in the field (Dohmen et al. 2016). A documentation of the more recent version 4.1.2 following ODD standards has not been published in the open literature but was available for this evaluation (Strauss 2017)[9].

C. crystallinus is a common predator in small fish-free ponds of Europe and feeds on small arthropods and rotifers (Berendonk and Bonsall 2002). Hatchlings develop in four aquatic larval instars before they pupate and emerge as adults for sexual reproduction. Females deposit an egg clutch on the water surface from which new larvae hatch. Under the temperate conditions used for model parameterization, *C. crystallinus* produces 2 - 3 generations per year and hibernates in a dormant stage in which larval development and growth are halted. Cannibalism in *C. crystallinus* has been reported to occur at high larval densities (Strauss et al. 2016).

The IBM *Chaoborus* population model is not spatially explicit but allows the simulation of a metapopulation by simultaneously simulating two separate populations that are connected by migration of adult midges. This way, recolonization in exposed habitats from non-exposed populations, as well as indirect effects on non-exposed populations via metapopulation dynamics may be addressed if migration rates can be reasonably estimated. The adult life stage is not explicitly modelled, emerged females are replaced with a fixed number of new eggs.

The model applies the concept of developmental rate summing: Newborn larvae start with a developmental state of 0 and emerge when this state has reached 1. Every day a specific increment is added to the developmental state that is specific for each sex and for each individual and increases with a combination of food availability and temperature. When the developmental state surpasses specific thresholds, an individual moves to the next life stage. During these transitions the larvae experience random background mortality. The only additional causes of mortality are cannibalism, larval dormancy in winter (fixed mortality rate), and pesticide exposure. Cannibalism acts only on 1st instar (L1) larvae; the daily probability of being preyed depends on the population density of 1st, 3rd (L3) and 4th (L4) instar larvae, but not of 2nd instar (L2) larvae. However, preyed L1 larvae do not contribute to the food supply (and thus the growth rate) of the preying larval instars. Food supply is modelled by default as constant pools for young (L1 and L2) and old (L3 and L4) larvae respectively. With increasing population size the available food per individual decreases, which prolongates larval development and thus the time window during which young larvae face the risk of being preyed. This way, survival of L1 larvae depends directly on the density of L1, L3 and L4 larvae and indirectly on the density of L1 and L2 larvae; this mechanism regulates the overall population density.

[8] Strauss, T., D. Kulkarni, T. G. Preuss and M. Hammers-Wirtz (2016): The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

[9] Strauss, T. (2017): Description of the individual-based model "IBM Chaoborus population model" for the aquatic phantom midge Chaoborus crystallinus. Research Institute for Ecosystem Analysis and Assessment (gaiac).

2.5.1.2 **Status of the Model**

The IBM *Chaoborus* population model is maintained by the gaiac Research Institute of RWTH Aachen University; the software and the code are not freely available. Publications on the model are mainly the work of Tido Strauss and colleagues, and the user community seems to be small. The model has been applied at least two times in dossiers for the risk assessment of pesticides (Tido Strauss, personal communication); see the case study in section 3.6 for an example. The first published version of the model (v. 3.9.5) included only the basic population model without a module for toxicant effects (Strauss et al. 2016). Predictions of this version on population dynamics during the breeding season without pesticide exposure were validated with mesocosm data (see below) following the pattern oriented modelling approach (Grimm et al. 2005). To be applied in risk assessment, the early versions of the model needed to be coupled with an external sub-model that provides effects of plant protection products at the individual level.

In the frame of the SETAC MODELINK workshop, Dohmen et al. (2016) applied the IBM *Chaoborus* population model (v. 3.0.9) to an anonymized, fast-acting and fast-dissipating pyrethroid ("modelmethrin") to demonstrate the applicability of the model for environmental risk assessment. In parallel to a mesocosm study, the model was applied to investigate whether Tier 1 individual-level effects are propagated to long-term effects at the population level. For this, the population model was coupled to a GUTS module (later included as built-in sub-model) that provided individual-level lethal effects. First, model predictions on long-term population effects of the pesticide were validated using the mesocosm data without further calibration. Then the model was applied to standard FOCUS scenarios, and the modelling results were compared to the conclusions drawn from a classical risk assessment approach. The model reproduced population recovery in the mesocosms reasonably well, though immigration was apparently overestimated (see details in section 2.5.2.9 and 2.5.2.10 of the model description below). No long-term effects on populations were observed in the mesocosms, and also the model predicted generally fast population recovery (within 8 weeks) for the mesocosms and the field scenarios. Long-term effects were only predicted for Northern scenarios characterized by only one reproduction peak per year due to low water temperature, and only if acute mortality reached up to 50 %. Therefore, the ability of the model to reproduce existing long-term effects has not been demonstrated yet.

A more recent version (v. 4.1.2, Strauss 2017) includes a classical EC50 approach and the GUTS module as built-in sub-models for individual-level effects. Additionally, mortality of dormant larvae during winter has been introduced. To reduce the synchronization of age-classes that is related to the model mechanisms, in v. 4.1.2 the density dependence of the food availability (intraspecific competition) can been switched off. Additionally, the normal distribution for the daily increment in larval development, from which larvae draw a random value at birth, has been replaced with a homogeneous distribution across ± 30 % of the mean value. Predictions on long-term population dynamics (overall larval population size and emergence of adults) of the new model version were validated with data from a three-year mesocosm study without pesticides.

The evaluation in this chapter is mainly based on the first published version 3.9.5 (Strauss et al. 2016) and on version 4.1.2 (Strauss 2017).

2.5.2 Model Description

2.5.2.1 Problem Definition

Context in which the model will be used

The IBM *Chaoborus* population model has been published by Strauss et al. (2016) to the scientific community to analyse the role of cannibalism on population dynamics of the phantom midge *Chaoborus crystallinus*. In risk assessment, the model may be used to extrapolate from pesticide effects at the individual level to effects on the population size and structure, to assess the time for population recovery after pesticide exposure, and to assess the risk of extinction under pesticide pressure.

Specification of the question(s) that should be answered with the model

No specific questions have been identified by the authors for the model as a whole, as these depend on the uncertainties that arise during the risk assessment of a specific substance or product. According to Dohmen et al. (2016), the model was developed for the extrapolation from toxic effects on individuals (assessed in the laboratory) to effects on populations in the field (usually assessed in semi-field experiments, i.e. micro-/mesocosms). Applications in case studies so far indicate that the model is primarily intended to assess the recovery time of population to meet control population size.

Specification of necessary model outputs and protection goals

For aquatic invertebrates, the protection goal is usually defined at the population level with either negligible effects (Ecological Threshold Option) or acceptable short- to mid-term effects (Ecological Recovery Option), e. g. effects of medium magnitude that last for up to 8 weeks on an edge-of-field scale (EFSA PPR 2010, EFSA PPR 2013). The required model output is therefore the change in size and structure of a contaminated population and a non-contaminated control population across days to months, from which the population recovery time can be estimated as the potentially most relevant endpoint for risk assessment.

Domain of applicability of the model

The model simulates the population dynamics of *Chaoborus crystallinus* in a seasonal (in terms of temperature and light regime) but otherwise stable aquatic environment with constant food supply and without interacting antagonistic species. The model has been developed for temperate ponds in Central Europe but might be parameterized also for different climatic conditions. The model proceeds in time steps of 1 d and can be run for many years. Custom exposure profiles may be entered, therefore the model may be coupled to $FOCUS_{SW}$ fate modelling. Without fundamental changes in the conceptual model, potential re-parameterization and application to other species is limited to taxa with similar life history, i. e. taxa with aquatic larvae that experience cannibalism and flying adults (e. g. some mosquitoes).

Why is the model being used?

Projecting from Tier 1 tests to effects in real freshwater populations is associated with uncertainty, as illustrated by the high assessment factors applied. Micro- and mesocosm studies aim to bridge the gap between controlled laboratory conditions and variable conditions in the field, but they are costly and can be therefore conducted only for one or very few scenarios. The IBM *Chaoborus* population model may be used to support Higher Tier studies by the projection from effects in Tier 1 tests to potential population effects in various scenarios.

The main purpose of the model is the prediction of population dynamics of the phantom midge, *C. crystallinus*, at different food levels and fluctuating temperature and light conditions in outdoor ponds based on individual life-cycles (Strauss et al. 2016). The specific question to which the model will be applied should depend on the uncertainties identified in the risk assessment of a specific substance or product. For an example see the case studies in the general information above.

What protection goal is being addressed?

The specific protection goal depends on a specific model application and cannot be addressed for the model in general. As outlined above, the typical protection goal to be addressed with the IBM *Chaoborus* population model is no long-term population decline beyond 8 weeks.

What outputs are required?

Addressing the potential protection goal outlined above will require the simulation of the size and structure of a population over time. Relevant endpoints that may be addressed with the model include the time for population recovery after exposure and the risk of populations becoming extinct.

How was the species chosen?

Chaoborus crystallinus is a common predator of zooplankton in fish-free standing freshwaters across Europe. No specific explanation has been provided on why the authors developed their model on this species. However, the species has been studied for many years, probably because it plays an important role in the freshwater macroinvertebrate community. Therefore, extensive literature is available for model development. In microcosms studies, larvae of chaoboridae turned out to be highly sensitive to insecticides (Heimbach 2000), which probably led to increased interest in the recovery potential of this taxonomic group.

Which other species/groups are being covered by the chosen one(s)?

Essentially, the model is quite specific for the genera *Chaoborus*. The model species may cover additional cannibalistic freshwater macroinvertebrates with similar migration potential as flying adults that are not more ecologically vulnerable or more sensitive to the studied pesticide than *C. crystallinus*. In the default parameterization for temperate conditions, populations seem to produce up to 3 – 4 generations per year (see e. g. the population peaks in Fig. 7). Thus, species or populations that produce less generations per year are not covered due to their higher vulnerability.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Model predictions of population dynamics without toxicant exposure have been tested using outdoor experiments under semi-field conditions (Strauss et al. 2016, Strauss 2017). In Dohmen et al. (2016), model predictions on population effects of PPP have been compared to a mesocosm study in which no long-term effects had been observed. See section 2.5.2.9 for details. The pattern matching between observed and predicted data has not been quantified and no degree of matching required for successful model validation has been pre-defined, probably because this has not been specified in any guidelines.

2.5.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

The authors collected and cited a large number of available studies on the life history of *Chaoborus crystallinus*. See section 2.5.2.7 for details. Additionally, the authors presented own experiments to investigate the effects of population density on the larval mortality (Strauss et al. 2016, Strauss 2017). From their experimental data the authors also developed a calibrated conversion scheme for the translation of measured densities of prey organisms to the food saturation level used in the model.

Assessment of the quality of the data

The literature cited and the additional experiments conducted by the authors of the model have been scientifically peer-reviewed. Most literature studies used for the parameterization of the model have been conducted in the laboratory under artificial conditions.

2.5.2.3 Conceptual Model

Description of the model concepts including a diagram.

The individual-based model simulates the life cycle of *Chaoborus crystallinus* with 4 larval stages, a pupal stage and the adult phase. Population dynamics emerge from the developmental rules for each individual that consider environmental conditions (food availability, seasonally changing temperature and photoperiod), competition for food among the larvae (optional), cannibalism of older larvae on first instar larvae, dormancy of larvae in winter, and migration of adults between populations. The adult life stage is not explicitly modelled; emerged females are replaced with new egg clutches or considered lost due to emigration.

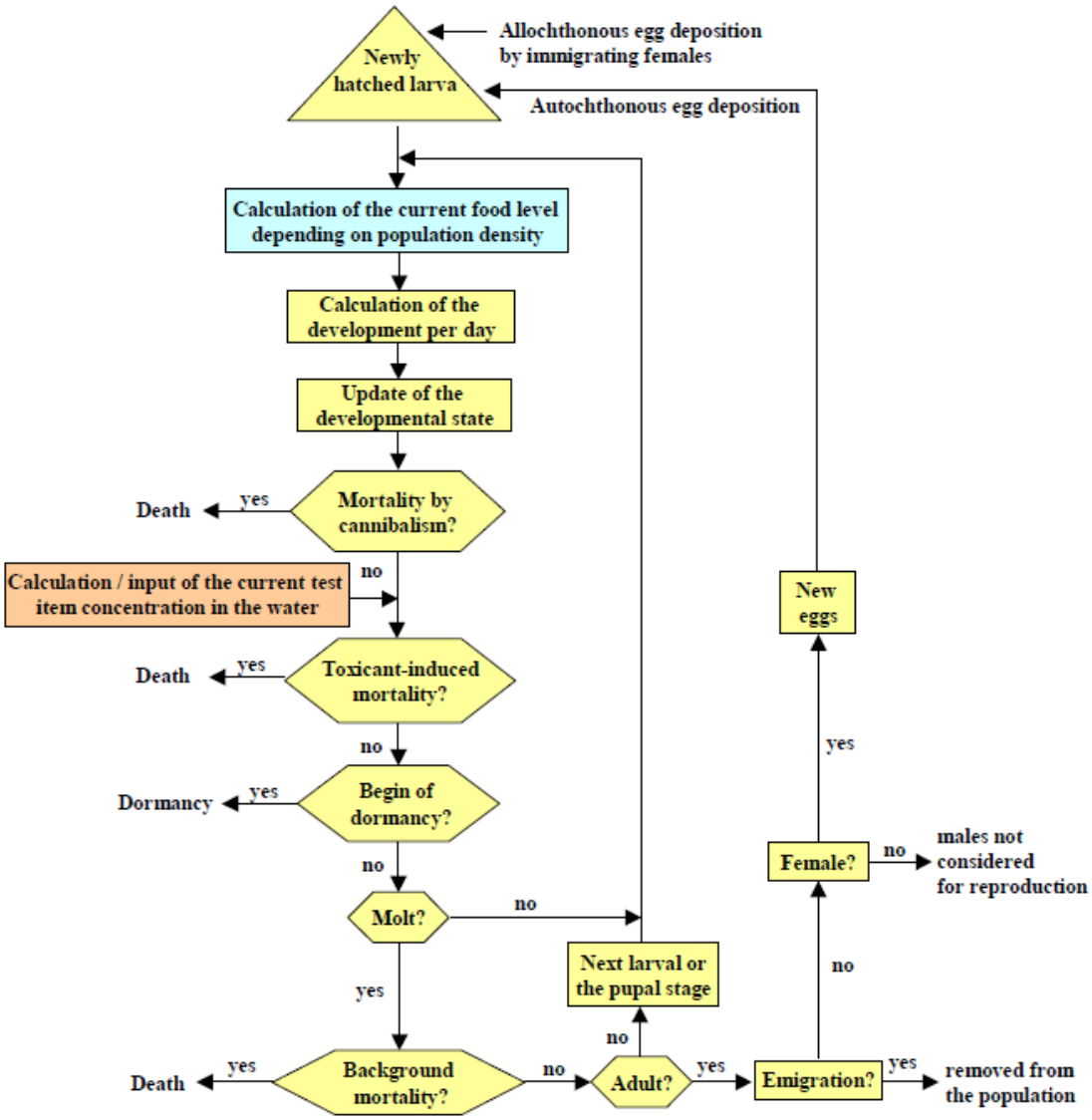
The overall food availability is considered constant throughout the simulation and is modelled as the ratio of the available amount of food relative to the amount of food required to obtain the maximum growth rate (nominal food saturation in percent). First and second instar (L1/L2) larvae share a common food source (that may represent e. g. rotifers), and third and fourth instar (L3/L4) larvae share a different food source (that may represent e. g. juvenile daphnids). The nominal food saturation levels are therefore scaled to the population density of L1/L2 or L3/L4 larvae to consider intraspecific competition, resulting in the actual food saturation (also in percent). The actual food saturation and temperature affect the daily developmental rate but not survival or fecundity. In version 4.1.2, the scaling of food saturation (effect of intraspecific competition) can be switched off to avoid the resulting synchronization of larval development.

Constant non-predatory background mortality is introduced at each moulting (transition from one life stage to the next). A combination of a critical photoperiod and temperature in autumn causes parts of the L4 larvae to become dormant and interrupt their development until spring. In v. 4.1.2 dormant larvae experience a temperature-dependent but otherwise fixed daily mortality. The model is not spatially explicit, i. e. chaoborids are assumed to be homogenously distributed in the water column. However, the model can simulate immigration and emigration of adults, using rates for the random addition or disappearance of emerged adults. Additionally, the model can consider metapopulation dynamics by connecting two simultaneously simulated populations via migration rates of adult individuals.

Probably the most prominent feature of this population model is the explicit consideration of cannibalism. Cannibalism is implemented as mortality for L1 larvae that increases with the density of L1 larvae (predation among same size class) and with the density of L3 and L4 larvae (predation between size classes). In simulation tests, the use of L2 larvae as predators has proven unrealistic, and population data have been less well met (Strauss, personal communication); therefore, the modellers assumed

that older larvae (L2 - L4) are not significantly preyed by conspecifics and that L2 larvae were not relevant as predators on L1. However, preyed L1 larvae do not contribute to the food saturation of any larval instar. The consumption rate of L3/L4 on L1 larvae is considered to increase exponentially with temperature. In contrast, no temperature-dependency of cannibalism among L1 larvae was implemented; the authors justified this approach stating that L1 larvae reach sufficiently high population densities only in summer, when temperature is close to the laboratory conditions of the experiments used for parameterization. Natural variability of life-history is incorporated by randomly sampling distributions for most life-history parameters (Monte Carlo approach).

Figure 6: IBM *Chaoborus* Population Model – Flowchart of the Conceptual Model



Summary of the main components and processes in IBM *Chaoborus* population model. Graph reproduced from Strauss (2017).

Identify the main components and processes in the system

The environment is represented by the nominal food saturation for young and old larvae, respectively, temperature, photoperiod, and water volume (converts population size to density). The population is represented by the summed up L1, L2, L3 and L4 larvae, pupae, adults and egg clutches. Individuals

differ in sex, the susceptibility to cannibalism during the first larval instar (random daily change), the susceptibility to background mortality during life stage transitions (random change at each life stage), and the susceptibility to dormancy in winter and to pesticides.

The main processes include larval development (increase in developmental state that determines life stage), background mortality during life stage transitions, mortality of L1 larvae from cannibalism, dormancy of L4 larvae in winter (interruption of larval development, and daily background mortality in version 4.1.2), emigration and immigration of adults, and effects of pesticides (if modelled).

How the effects of the chemicals are modelled

The basic population model published in Strauss et al. (2016) needs to be coupled to an external module for the simulation of individual-level pesticide effects. In version 4.1.2, two modules for individual-level effects have been incorporated (Strauss 2017). This includes an EC50 approach meant for fast-acting, rapidly degrading pesticides, and a GUTS module for longer-lasting compounds. In the applications published so far, only acute mortality has been implemented. Daily changing pesticide concentrations can be either entered from separate fate modelling, or can be calculated within the model from user-defined exposure events and dissipation following first-order kinetics.

How the components and processes are linked

Each time step (24 h), day length and water temperature are calculated from internal equations if no measured data are imposed. Based on the current larval densities, the actual food saturation levels are updated (optional in version 4.1.2). Then the developmental state of each individual is increased by a rate that increases with temperature and the actual food saturation. Background mortality during life stage transitions is fixed. The daily mortality of L1 larvae from cannibalism increases with the density of L1 larvae and of L3/L4 larvae, respectively. The effect of L3/L4 larvae increases with temperature.

As a consequence, increasing nominal food saturation or temperature will accelerate the larval development (and thus shorten the generation time and potentially accelerate population recovery from pesticide effects). In contrast, increasing food saturation or temperature may have only moderate effects on the carrying capacity (population size) in the model (see also the sensitivity analysis in section 2.5.2.8): Accelerated larval development shortens the period when larvae are exposed to predation, but is expected to be compensated by an increased predation risk during that critical period because predation increases with population density and temperature. This way, different settings of food saturation and temperature may produce relatively similar control runs but different times for population recovery after pesticide effects. Food saturation levels and the temperature profile should be therefore particularly considered in the evaluation of an environmental scenario for risk assessment.

If the combination of water temperature and photoperiod falls below an individual threshold, L4 larvae become dormant. Dormancy halts larval development and, in version 4.1.2, induces daily background mortality that increases with temperature. Dormancy ends in spring when temperature increases above a threshold (not affected by photoperiod). Reproduction in the model is not affected from the environment. When an adult female emerges, either a fixed number of new eggs are immediately added to the population or, with a user-defined probability, the individual is considered lost due to emigration. A user-defined proportion of emigrated adults from one population can be added to the adults from a second population (immigration). Pesticide effects do not interact with other simulated processes at the individual level.

2.5.2.4 **Formal Model**

Identification of the model variables

Altogether, 14 variables were identified that describe the environment and population, and 7 variables for the individuals. The environment is represented by the three driving variables water temperature, photoperiod and pesticide concentration. Daily values for these variables can be provided as input data or calculated from internal equations. Additionally, the environment is represented by two state variables for the actual food saturation of L1/L2 and L3/L4 larvae, respectively. The population is represented by nine state variables that describe the number of individuals per life stage (eggs, L1 - L4 larvae, pupae and emerged adults) and the number of deceived individuals.

Each individual is characterised by two state variables: dormancy and the developmental state. The developmental state determines the life stage, using the "concept of developmental rate summing" (Strauss et al. 2016): New born larvae start with a developmental state of zero; the developmental state increases every day depending on the food saturation and water temperature. Maturity is reached if the developmental state reaches a value of 1. Other thresholds for the increment in developmental state indicate the transitions from one larval instar to the next (0.211, 0.378, 0.548) and to the pupal stage (0.860). Additionally, each individual is characterized by its susceptibility to cannibalism as L1 larva (probability of being preyed, randomly renewed from a probability function every day), and a susceptibility to background mortality during life stage transitions (randomly renewed in every life stage). Finally, each individual is characterized by its sex, its susceptibility to dormancy and its susceptibility to pesticide effects, all of which are drawn from random functions at birth.

State variables and properties have been listed in Table S.1 of Strauss (2017). Please note that in the model documentation, the individual susceptibilities have not been listed. The listed pond volume has been considered as parameter (see next section) because it seems not subject to change in the course of a simulation according to the model descriptions.

Identification of the model parameters

The environment is characterized by 6 – 17 parameters that may be set up for a given environmental scenario: The water volume translates the population size to population density. The nominal food saturation levels for L1/L2 and L3/L4 larvae, respectively, determine (together with temperature) the maximum growth rate that can be achieved in the simulations. For repeated simulation runs, food saturations can be initialised using random values drawn between a user-specified minimum and maximum value. To set up the nominal food saturations, conversion schemes have been developed that relate food saturation to a given density of prey organisms that may be assumed or observed in a study. This conversion requires two additional parameters (one for each food source) that are not directly part of the simulations. If no external data is provided, the daily water temperature and photoperiod are calculated from sinusoidal functions that require three parameters for temperature and two parameters for the day length. Additionally, the dates of pesticide application, the initial concentration in the water and the half-life period (DT50) of the pesticide is required if no daily exposure profile is provided. Depending on the applied sub-models for individual-level effects (LC50 or GUTS), 2 – 5 additional parameters on pesticide properties are required (included in the overall number of parameters stated above).

Additionally, the model comprises a set of 27 – 29 basic biological parameters with a built-in parameterization that may not need to be changed as long as the model is applied within its domain of applicability: Two half saturation constants are required for the scaling of nominal to current food saturations (may be excluded in version 4.1.2). Four parameters determine the daily progress in the developmental state of individuals and its dependency on food saturation, temperature and sex. Six thresh-

olds determine at which developmental state individuals move from one life stage to the next. Four parameters are used to scale cannibalism rates of L1 and L3/L4 larvae depending on population density and temperature. An additional parameter determines the individual susceptibility of L1 larvae to mortality from cannibalism (renewed each day). Three parameters determine the individual susceptibility to background mortality during life stage transitions (renewed after each moulting), the probability of loss of adults due to emigration and the probability of an emigrated adult to immigrate in the second simulated population. Another three parameters determine the probability that an emerged female produces an egg clutch, the clutch size, and the sex ratio at birth. Additionally, six parameters are used for the scaling of dormancy depending on photoperiod and temperature, and of mortality during dormancy depending on temperature.

Description of the most important model equations or algorithms

Processes in the IBM *Chaoborus* population model are mainly based on equations rather than rules. Population dynamics and the related endpoints of potential interest for risk assessment (such as recovery time) emerge from the interactions of at least 20 equations identified from the different submodels. The most important equations for the calculation of the variables described above are presented in Tab. S.3 of the model documentation (Strauss 2017).

Water temperature and day length (if not user-defined) follow a sinusoidal function. The effects of food saturation and temperature on the daily developmental rate are modelled to be linear and additive. The actual food saturation decreases with increasing larval density following a Michaelis-Menten equation (if not switched off). Cannibalism rates for L3/L4 larvae are modelled to increase exponentially with temperature.

2.5.2.5 Computer Model

Description of the model implementation

The model was implemented in Delphi®Professional 5.0.2.1.3. The code or a detailed description of the model implementation is not publicly available.

Checking the computer model for errors, bugs and inconsistencies in the code

No systematic procedure has been described in the available publications.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

No specific verification of the software implementation has been published. However, during validation the model apparently performed as expected and reproduced various patterns of population dynamics in non-exposed populations (see section 2.5.2.9).

2.5.2.6 The environmental scenario

Description of the environmental scenarios, i.e. the environmental context in which the model is run.

The environmental scenario is specific for each model application and cannot be described in general. For validation purposes, Strauss et al. (2016) applied the model to two outdoor microcosm experiments without pesticide exposure (see section 2.5.2.9). The parameter setting used in both applications may be considered as a default environmental control scenario that was also used for a sensitivity analysis (see section 2.5.2.8).

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

In the default scenario mentioned above, model parameters were matched to the environmental conditions of the microcosm performed under temperate conditions (see section 2.5.2.10). Temperature and photoperiod were set to match conditions in small ponds at 51° latitude in Central Europe. This scenario is probably representative for fish-free ponds in the EU Central Zone.

2.5.2.7 Parameter Estimation

Description of the model parameter estimation

Strauss et al. (2016) performed several experiments for the parameterization of cannibalism. Predation rates of L4 larvae on L1 larvae as well as the predation of L1 larvae amongst themselves were studied at 20°C in the laboratory. According to the authors, L3 larvae also showed considerable predation in pre-tests that have not been described in detail; despite the difference in size, they were therefore considered to prey similarly on L1 as L4 larvae. All experiments were conducted in filtered medium that contained no food for the *Charoborus* larvae. The larvae were obtained from an outdoor pond. To address cannibalism among L1 larvae, L1 were grown in the laboratory at densities ranging from 3 to 200 ind./L, and the mortality was recorded after 48 h. The increase in mortality with larval density followed a saturation curve. To address cannibalism of L4 on L1, L1 larvae with population sizes ranging from 2 – 50 individuals were grown in presence of a single L4 larva, and the mortality of L1 was recorded after 48 h. The test design was chosen such that the mortality of L1 did not exceed 25% to avoid an underestimation of predatory rates owing to excessive reduction of the L1 larval density during the tests.

Additionally, prey selectivity of L4 larvae was tested in a lab microcosm stocked with 30 L4, 60 L1 and two density levels of alternative prey species (cladocerans, copepods) from outdoor ponds. The change in the community composition after 72 h was recorded. Results showed that L4 larvae generally favoured daphnids and that preference for L1 larvae over alternative prey did not change with food levels.

Parameters estimated from the literature — what are the sources and why are these appropriate?

Parameters from the literature were obtained from peer-reviewed scientific publications and from academic theses. E. g., the developmental rates of larvae and their dependency on temperature and food, as well as the background mortality during moulting were taken from laboratory studies with individually grown larvae (Büns and Ratte 1991, Niewersch 2005). The sex ratio at birth and the clutch size were taken from outdoor mesocosm studies (Havertz 1988, Sevim 2012). The temperature-dependency of L3/L4 cannibalism was established from a study on food webs in lakes (Ramcharan et al. 2001). The dependency of non-cannibalistic ingestion on prey density was parameterized using a laboratory study on L2 larvae of *C. punctipennis* feeding on rotifers (Moore 1988), and a laboratory study on L4 larvae of *C. crystallinus* feeding on *D. magna* (Wiertz 1984). The loss of emigrating adults (net emigration rate) was calculated from a semi-field study that compared oviposition with and without mesh cover (Sevim 2012).

Parameters obtained from calibration — how and why this was done?

The food saturation levels depend on the assumed (or observed) food quantities and have to be adapted to each specific scenario. Strauss et al. (2016) developed a conversion scheme from zooplankton densities to food saturation levels by calibrating the model to two outdoor experiments that were conducted mainly for model validation (see section 2.5.2.9). In the first experiment, larvae hatched from 200 – 240 eggs and developed under high or low food conditions in 10 L buckets. Under high

food conditions, the larvae were provided with a specific amount of water from a eutrophic fish pond every 2 – 4 days, which contained on average 371 cladocerans (considered as food for L3/L4 larvae) and 2,714 rotifers and nauplii (considered as food for L1/L2 larvae) per litre. In the second experiment, a *C. crystallin*us population was continuously cultured in covered mesocosms and provided with 60 cladocerans and 115 nauplii and rotifers on average. The model was then calibrated to these experiments. The exact procedure has not been described, probably the food saturation level was tweaked such that the predicted time for larval development matched the observations. The resulting relations of food saturation vs. prey density matched the functional responses described in the studies of Wiertz (1984) and Moore (1988) (see above). All these data together were used to establish a conversion scheme used for future applications (Strauss et al. 2016).

2.5.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

A local sensitivity analysis was conducted for the model version 3.9.5 that addresses the model sensitivity relative to the point estimates chosen and not for the entire parameter distribution. It is thus valid for the specific scenario analysed that may be considered the default scenario for model demonstration purposes (see section 2.5.2.6). The sensitivity of five endpoints (mean larval abundance, larval abundance at the end of the simulation, sum of emerged adults, sum of eggs and sum of dead larvae) to 11 potentially relevant parameters were tested. Fifty Monte Carlo simulations were conducted with a fixed starting density of 4,000 larvae and a constant food level of 25%. For each simulation run, one of the selected parameters decreased and increased by 10 %. Sensitivity of the population recovery time after an acute decrease in population size, which is probably the most interesting endpoint of this model for risk assessment, has not been analysed.

The number of emerged adults was highly sensitive to the temperature-dependency of the larval developmental rate, and moderately sensitive to the food saturation level of L3/L4 instars and to the general larval developmental rate. Cannibalism did not considerably affect emergence success, but moderately decreased the larval abundance. However, larval abundance seemed generally less sensitive than emergence, probably due to buffering by the density dependent processes in the model. The authors concluded that the model is a buffered system that is not forced by a single process; therefore, the model is relatively insensitive to small changes in single processes or parameters.

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

An extensive uncertainty analysis is case-specific, therefore most appropriate once the model has been set up for a specific pesticide and protection goal. As a rule, parameters that show high impact in the sensitivity analysis on endpoints relevant in risk assessment have the potential to contribute significantly to the uncertainty of the model. In the described model applications, uncertainty in model predictions is indicated by the 95 % confidence intervals from repeated model runs, but the contribution of different sources to this uncertainty has not been investigated.

2.5.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

Strauss et al. (2016) performed two outdoor experiments to test the ability of the model v.3.9.5 to reproduce various patterns of the population dynamics without pesticide exposure. In the first study,

fresh clutches of *C. crystallinus* with 200–240 eggs each were introduced into 10 L microcosms and covered by nets to avoid dispersal of adults. Every 2–4 days, the larvae that emerged from the eggs were fed with a high or low amount of zooplankton and the development of the age class cohorts was recorded. In the second study, *C. crystallinus* populations with a mixed population structure developed in 4 m^3 covered mesocosms from June to November. The adults copulated and oviposited under the nets, which enabled the studying of the full life cycle of at least one generation. The food saturation levels were calibrated to these studies (see section 2.5.2.7) and 50 simulations were run for each of the scenarios. Additionally, with the data from the second experiment, the relevance of different model mechanisms has been tested by switching on and off the background mortality rate or cannibalism among L1 or by L3/L4 larvae.

The ability of the model v. 4.1.2 to predict long-term population dynamics without pesticide exposure across the seasons has been tested using data from eight artificial outdoor ponds over three years (Strauss 2017). However, the same data have been used to calibrate the mortality of dormant L4 larvae, therefore only the population development in summer can be considered as subject to validation with really independent data. Each microcosm was characterized by 4 m^3 water, a sediment layer of 10 cm and aerobic conditions throughout the experiment, and was probably covered by a mesh. Environmental conditions in the model were matched to the experiment, incl. regularly measured water temperatures.

Additionally, Dohmen et al. (2016) tested predictions of the model v. 3.0.9 on the population recovery after pesticide exposure with data from a mesocosm study on the anonymised pyrethroid "model-methrin" (Fig. 8). The mesocosm study comprised a series of 3 pesticide applications with a 7 d interval and lasted for 50 d after the first exposure. The overall larval population size was assessed 1, 2, 3, 6, 9, 22 and 44 days after exposure. With nominal pesticide pulses of 0, 10, 25, and 50 ng/L, the acute mortality in exposed mesocosms ranged from ~85 % to 100 % six days after the first exposure. The environmental parameters of the IBM *Chaoborus* population model were matched to the mesocosm study, but no further calibration was done for the built-in parameterization of the basic biological parameters (see section 2.5.2.7). The population model was coupled to a GUTS module that simulated the individual-level effects (only mortality). While the population model proceeded in daily time steps, hourly time steps were used for the GUTS module. GUTS was calibrated with laboratory data (survival over time within 96 h constant exposure). However, in the mesocosm study rapid dissipation of the pesticide was observed and accordingly simulated. Therefore, first the ability of the parameterized GUTS module to predict effects of different exposure profiles was validated, using independent data on short-term mortality observed in the mesocosms within six days after first pulse exposure. Then the population model was validated using data on long-term population development in the mesocosms. Results and conclusions from the modelling and mesocosm study were compared (see next section).

The descriptions of the validation studies suggest that the experimental populations of *C. crystallinus* were not embedded in a (close to) natural community but experienced limited or no interaction with antagonistic species such as amphibians and large invertebrates who may act as predators and competitors.

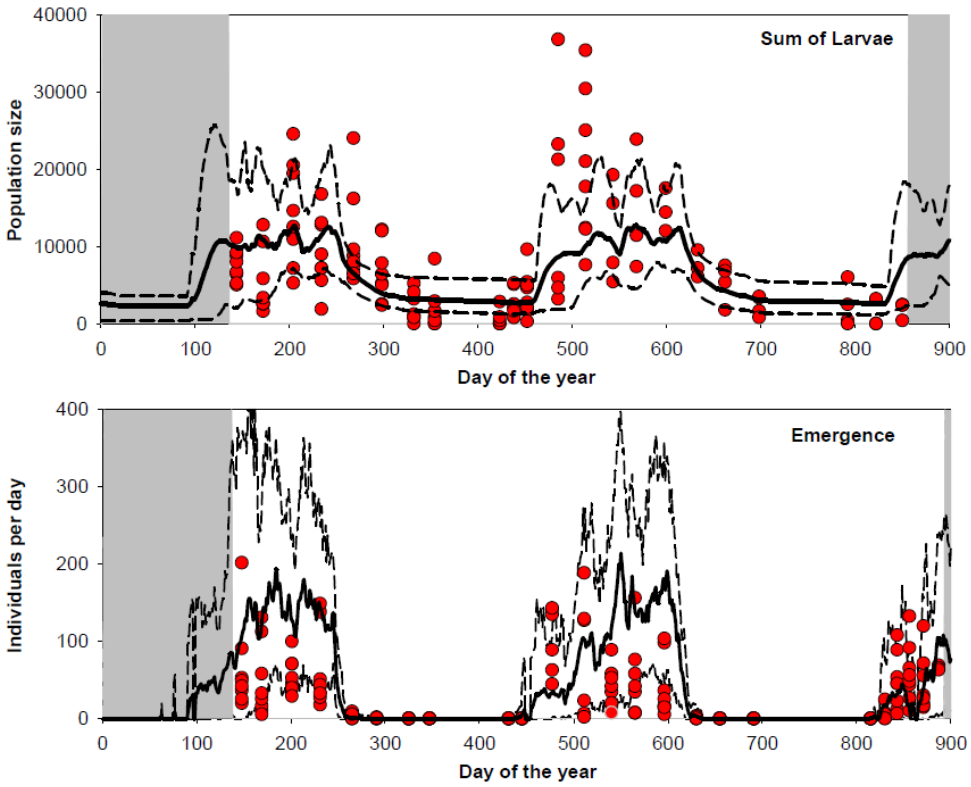
Demonstration that the model output provides an adequate match to data patterns

In Strauss et al. (2016), the IBM *Chaoborus* population model predicted the dynamics of the mean overall larval population size and of the different larval stages in the first experiment quite well, both under high and low food conditions. However, under high food conditions, development across the larval instars seemed to occur a bit faster than predicted by the model. The observed variation in the abundance of each instar across experimental replicates was considerably higher than the range of predictions from the 50 simulation runs, probably because the variation in the simulated initial L1 density was considerably lower than the observed variation. Therefore, the predicted variability in abundance cannot be validated with the observed variability.

The model also predicted the population dynamics in the second experiment from Strauss et al. (2016) reasonably well, though the peak of L2 occurrence was predicted too early in the model. Switching off background mortality during moulting or cannibalism did not considerably affect pattern matching of model outcomes with observed data, except for the removal of cannibalism by L3/L4. Therefore, the mortality during moulting or from predation among L1 larvae was compensated by other mechanisms of density-regulation (competition). In this experiment, the predicted and observed variation in the number of individuals matched reasonably well, probably because the initial overall population size was varied randomly between 1,500 and 4,800 ind./L which better matched the observed initial population size than in the first experiment. The mean and the variation in the number of emerged adults appears to be underestimated by the model. However, the model predicts emergence on a daily basis, whereas adults were counted weekly in the experiment. Because adult of *C. crystallinus* live for up to six days, it may be more appropriate to divide the observed numbers of adults by the average life span (probably four days) before matching to model predictions. It is not clear whether this was done. If not, model predictions on emergence might be better than suggested by the presentation in Strauss et al. (2016) in summer, but overestimated emergence in autumn.

Long-term simulations of the model version 4.1.2 over three years in Strauss (2017) showed that the predicted timing of adult emergence and the following increase in population size due to reproduction matched the experimental observations well (Fig. 7). During the breeding season, the overall larval population size observed in the artificial ponds showed high variability but generally matched the model predictions. Data on the population structure have not been presented. The average number of emerged adults during the breeding season seemed to be overestimated by a factor of approx. 2 in the model. The decrease in population size during winter was calibrated to the data and therefore cannot be considered as part of the validation.

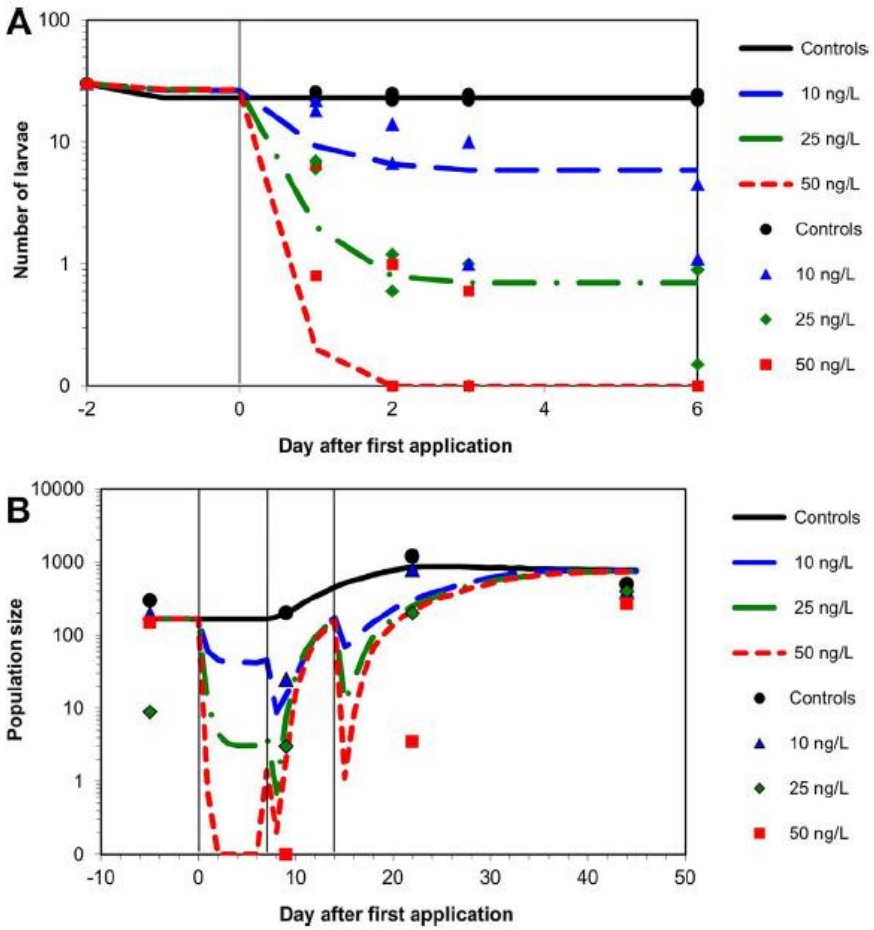
Figure 7: IBM *Chaoborus* Population Model – Validation of Long-Term Population Growth



Comparison of population dynamics predicted by the IBM *Chaoborus* population model v. 4.1.2 and observed in artificial outdoor ponds over three years. Lines represent the mean, minimum and maximum of model simulations, red dots represent experimental observations. Top: sum of all larvae; Bottom: number of hatched adults per day. Graphs reproduced from Strauss (2017).

Dohmen et al. (2016) tested the ability of the model to predict recovery times after pesticide exposure. First, the ability of the GUTS module to predict acute individual-level effects in the mesocosms was validated. The predicted and observed mortality during the first 6 days after the first pulse exposure were compared. GUTS predicted mortality reasonably well, thought it tended to overestimate acute mortality after 24 h and to underestimate delayed mortality after 6 days: In the mesocosms, individuals continued to die from day 3 to 6 when the pesticide was dissipated, which was not predicted by GUTS (Fig. 8a). Afterwards, the ability of the whole population model to predict population recovery in the mesocosms was tested. Predicted and observed population sizes at 9, 22 and 44 d after the first exposure matched reasonably well, except for the number of larvae 22 days after exposure to the highest concentration (50 ng/L) which was considerably overestimated (Fig. 8b). This concentration caused larval populations to become extinct both in the model and in the mesocosms, and with immigration switched off the populations remained extinct in the model. The results suggest that recolonization was a relevant process in the mesocosms, particularly after local extinction, but that the immigration rate was set too high in the model. In contrast, pattern matching at lower concentrations suggests that autochthonous recovery from reproduction was predicted well. However, long-term effects (> 3 weeks) have neither been observed in the mesocosm nor in the model, therefore conclusions that can be drawn on the ability of the model to reproduce such long-term effects remain very limited.

Figure 8: IBM *Chaoborus* Population Model – Validation of Recovery after Pesticide Exposure



Validation of model predictions for the effects of the anonymised pyrethroid "modelmethrin" using data on short-term effects (a, up to 6 d after exposure) and long-term effects (bottom, up to 30 d after the last exposure) from a mesocosm study. The upper panel (a) is testing the ability of the GUTS module to predict individual-level effects in the mesocosm study with pulse exposure after calibration to laboratory data with constant exposure. The lower panel (b) validates the ability of the whole population model to predict population recovery after 3 pulses of pesticide exposure. Dots represent observed values, lines represent model predictions. Grey vertical lines indicate pesticide applications. Graphs reproduced from Dohmen et al. (2016).

2.5.2.10 **Model Use**

Explanation of how the model conforms to the requirements set in the problem definition.

The authors considered the model suitable for regulatory purposes, noting that it reproduced various patterns of population dynamics well, including the recovery after pesticide exposure (Dohmen et al. 2016, Strauss et al. 2016).

Description how the model works (user manual).

A user manual is not publicly available, but the model has been documented in detail in Strauss et al. (2016) and in Strauss (2017) following the ODD guidelines (Grimm et al. 2006).

Description of the pesticide parameters values used in the model.

Parameter values must be fitted case-specific for each model application.

Description of the specific assessment including a discussion of the most important results.

This item cannot be addressed for the model in general, but only for a specific application. As an example, after model validation Dohmen et al. (2016) applied the IBM *Chaoborus* model to field scenarios with the same pesticide in order to demonstrate potential use of the model in risk assessment. The model was applied to two standard FOCUS scenarios for ditches (D1 and D3) and ponds (D4 and D5), respectively, each with a series of 5 pesticide applications. In the ditch scenarios, a control and a treatment simulation were run which were coupled by 50 % exchange of adults to consider metapopulation dynamics. This way, the control scenarios were not fully independent but could be indirectly affected by the pesticide. This approach impedes the assessment of effect sizes from the comparison of population dynamics in the treated and the control scenario. 50 simulations were run per scenario; each run, the food saturation level was randomly selected from a range between 10-50 % of the maximum uptake rate. Due to stochasticity and the variation in the food saturation level, the maximum deviation between 2 independent simulated control populations (daily average of the mean value for 50 single runs) was up to 25%. Therefore, the authors considered only pesticide effects as adverse when they reduced the average population size by more than 30 %.

Overall, the model predicted fast population recovery from modelmethrin effects (within 8 weeks) due to reproduction and recolonization, even when each pesticide application reduced the population by approx. 30 %. In non-exposed northern ditches (D1) the IBM *Chaoborus* population model predicted a single peak in population size over the year. In southern ditches (D3), the higher water temperature resulted in an extended reproduction period with two peaks in population size. Therefore, the predicted risk of modelmethrin was larger in northern scenarios, where the pesticide was applied later in the year, while the reproduction season, during which population recovery was possible, was shorter. Accordingly, when a single pesticide application killed approx. 50 % of the population, long-term effects were predicted for northern, but not southern ditches. In the pond scenarios, the pesticide concentrations were lower but exposure lasted longer than in the ditches. Long-term effects occurred only when ponds were considered to have no buffer zone (increased exposure) and when migration between ponds was switched off (Dohmen et al. 2016).

2.5.3 Model Evaluation

2.5.3.1 Problem Definition

The regulatory context in which the model is run

The IBM *Chaoborus* population model has been developed with the aim of supporting Higher Tier studies on freshwater invertebrates by extrapolating toxic effects from individuals in the laboratory to populations in the field. In most mesocosm studies only a single environmental scenario can be tested in which recolonization is likely to differ from scenarios in the field. With simulation studies, the risk assessment may be extended to a range of different scenarios, including varying rates of recolonization. In this context, the population model may be potentially applied to assess the risk of long-term population effects, i. e. the time to population recovery or the probability of population extinction.

The question that has to be answered with the model

Based on the published model applications, the model seems mainly intended for the studying of population recovery after pesticide exposure. Predictions of the IBM *Chaoborus* population model on recovery times are associated with considerable uncertainty (see below); therefore, potential model applications should follow a probabilistic approach. This approach may not focus on the demonstration that recovery within a short time will occur in the model for a specific scenario. Instead, the application may focus on the margin of safety that the recovery time will not exceed a threshold even when parameters and processes such as recolonization are modified to meet various (realistic or non-realistic) worst-case assumptions.

The available knowledge and data relevant to the risk assessment question

The authors of the model have presented a thorough overview on the knowledge available on the life history of the model species. Studies presented by the authors show that recovery of *C. crystallinus* after pesticide exposure through reproduction and recolonization may be fast and a relevant aspect to be considered in risk assessment. Therefore, the publications discussed the applicability of the model to the specific protection goals in risk, though the representativity of the model species as worst case and thus the applicability to the actual protection goals concerning invertebrate abundance and diversity (EFSA PPR 2013) is unclear. There is substantial uncertainty on how additional mechanisms may affect population recovery that have not been incorporated in the model, such as interactions of life history with individual-level sensitivity or indirect effects from interaction with less sensitive species.

The model simulates a (meta)population that is isolated from other species. However, populations in mesocosms and particularly in the field may experience interactions with antagonistic species. E. g., *C. crystallinus* may experience competition from notonectidae, copepoda, zygoptera and amphibians that also feed on cladocerans and/or rotifera (Sprules 1972, Bohle 1995, Van de Meutter et al. 2005, Hamilton et al. 2012). Interactions with antagonistic species may considerably affect the recovery potential of populations (Liess et al. 2013). E. g., recolonization after local extinction has been overestimated in the simulations of Dohmen et al. (2016; see validation study in the model description above). This may have happened because in the absence of *C. crystallinus* following high exposure, less sensitive competitors might have occupied the common ecological niche (Foit et al. 2012). Therefore, conclusions from the modelling on populations in the field should be drawn with care, and uncertainties resulting from the limitations outlined above should be thoroughly communicated.

The outputs required to answer these questions including performance criteria for the regulatory model

The required output depends on the specific protection goal addressed. To address the risk of extinction and the time to recovery, predictions on the population density (and possible population structure) are required for an extended period of time after exposure, including confidence intervals that propagate uncertainty in the model input to uncertainty in the model predictions. The IBM *Chaoborus* population model can provide such endpoints.

The species to be modelled

The phantom midge *Chaoborus crystallinus* represents a common macroinvertebrate in fish-free ponds. It is a relevant predator of zooplankton species such as water fleas and rotifers. The related species *C. flavicans* in lakes is an important food source for young fish. Chaoboridae are therefore of high ecological relevance in food webs and they are highly sensitive to various insecticides, which justifies particular interest in this model species.

However, the actual protection goal of Higher Tier studies is the whole community. As the model is very detailed in the simulation of a single model species, generalizing the modelling results to other freshwater macroinvertebrates is difficult. E. g., simulation of cannibalism is not applicable to non-predatory taxa, and predicted recovery times for *C. crystallinus* may not be protective for species with longer generation times or lower reproduction rates. The generation time of *C. crystallinus* is short (2 - 3 generations per year in Central Europe) and thus the recovery potential is high as compared to univoltine species. Therefore, the modelling should be supplemented with additional studies on other species.

Requirements for the environmental scenarios to be used in the risk assessment

To apply the model, case-specific information on the modelled water volume, the initial population size and structure, the recolonization rate, pesticide application and on the daily water temperature, photoperiod and food supply are required. Therefore, uncertainties in characterizing an appropriate environmental scenario will exist but are probably not larger than for the FOCUS scenarios used in fate modelling. A conversion scheme has been established from laboratory experiments to convert the assumed food supply into a food saturation level required by the model, considering the maximal feeding rate of the model organism. However, feeding rates may be lower in the field e. g. due to predator avoidance behaviour or the need of searching food in a complex environment; therefore, the conversion scheme may be used with caution.

2.5.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

The development of the model is based on a comprehensive literature review on the life history of *C. crystallinus*. Additionally, specific experiments have been performed to parameterize the model. However, the biological part of the population model (e. g. growth rates) has been mostly parameterized with data from laboratory studies. Therefore, potential additional environmental stressors that are not explicitly modelled (e. g. low oxygen levels, temperature stress and increased density stress due to partial desiccation of ponds during summer, or competing and predacious species such as water scorpions and tadpoles) have also not been incorporated implicitly via the use of parameter values from (close-to) field observations where such stressors may exist. The model thus simulates a population under laboratory conditions which are typically considered optimum, except of the explicitly modelled environmental factors (food supply, temperature, population density, photoperiod). This is not fully reflecting field situations and needs to be considered when making predictions for real populations.

Has the quality of the data used been considered and documented?

Information on the model species has been has been adequately referenced and obtained from peer-reviewed scientific publications and doctoral theses which. The experiments performed by the authors have been published in a peer-reviewed journal and seem to be of high quality.

Have all available data been used? If not, is there a justification why this information has not been used?

No statement has been provided regarding a possible selection of available data used. Given that ecological modelling generally has to deal with a lack of data and we are not aware of relevant unused data, it seems plausible that all available data have been used.

2.5.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

The model predicts the numbers of individuals for each developmental state in daily time steps over a user-specific time span. The model was therefore designed such that the protection goals of no long-term decline, and of population recovery within up to eight weeks for freshwater macroinvertebrates (EFSA PPR 2010, EFSA PPR 2013) can be addressed.

Are the modelling endpoints relevant to the specific protection goal?

The model predicts various endpoints such as the population density of all life stages, and also the number of individuals that have died or emerged. With Monte Carlo simulations, the daily arithmetic mean, minimum and maximum values are calculated from these population properties and saved for each simulation run. With these data, protection goals for aquatic invertebrates such as the magnitude of long-term decrease, time to recovery and the extinction risk can be addressed. However, the results address a single model species and have to be extrapolated to the whole community under field conditions as the ultimate subject to be protected, which requires additional information.

Is the modelling approach justified?

The IBM *Chaoborus* population model follows well established principles of individual-based modelling which is an appropriate approach because it can consider individual heterogeneity in sensitivity and life history traits. Compared to other IBMs such as MASTEP and eVole, mechanisms in this model are mainly described by a set of connected equations rather than by behavioural rules for individuals. The model is not spatially explicit and thus assumes that spatial heterogeneity in exposure may be neglected. This may be justified at the spatial scale of edge-of field standing surface waters. However, effects of pesticides on real populations may be modified by the community context which is beyond the scope of a pure population model.

Is the conceptual model logical?

The model concept is generally logical. In the model, food limitation delays the larval development but does not directly increase mortality. Starvation to death was probably not implemented because starving larvae may compensate food limitation by feeding on their conspecifics. Indeed, cannibalism on early larval instars has been considered an important mechanism that enables predatory freshwater macroinvertebrates to survive when food for late instars is low or when the timespan suitable for development is short (Wissinger et al. 2004, Persson and de Roos 2006).

However, potential inconsistencies regarding the implementation of cannibalism have been identified: Cannibalism as an intraspecific form of predation is expected to be a +/- relationship, i. e. beneficial for the predator and adverse for the prey. In the model, cannibalism increases mortality of preyed L1 larvae but provides no benefit for preying larvae, such as increased food saturation (and hence development) or decreased background mortality. Therefore, cannibalism is modelled as a is 0/- relationship. Such a relationship represents density stress on young larvae (e. g. asymmetric competition for space or physiological stress that affects predominantly L1 larvae) rather than cannibalism. Implementing effects of cannibalism not only for the prey but also for the predators may have consequences for the simulated population recovery from pesticides: Pesticides that kill a high proportion of L1 larvae (which is typically the most sensitive life stage) may eliminate an important food source for the older larvae. Indirect effects of pesticides on older larvae via the food chain are therefore to be expected, which may significantly delay population recovery. However, according to the author of the model, the energy gain from cannibalism may be noticeable for small larvae under certain conditions, but it is negligible compared to the turnover of further prey (personal communication). This assumption could be checked in the future with a DEB model variant for *Chaoborus*.

Additionally, it seems unusual that L1 larvae are sensitive to predation by all but L2 larvae in the model. Exclusion of L2 as predators improved the matching of predicted and observed population data (Strauss, personal communication) but the reason has not been explained. L4 (and L3) larvae showed a significant impact on L1 survival in the experiments for parameterization, probably due to the difference in size. L1 larvae were considered relevant as predators because they can reach high population densities. Therefore, it seems logical to consider L2 larvae to contribute to the predation-relevant density of small larvae at least in the same way as L1, specifically since they are considered to share the same food source in the model. Finally, the assumption that no relevant predation occurs among older larvae should be experimentally justified; e. g., in the mosquito *Culex pipiens* it was observed that L4 prefer L3 larvae (larger prey) over early larval instars when preying on conspecifics (El Husseiny et al. 2018). Incorporating additional cannibalism between more larval instars will increase larval mortality and thus may lower the potential for population recovery. However, the predicted population capacity that emerges from density-dependent mortality in the model, has met observations, so that the overall larval mortality seems generally adequate.

Are the processes included in the model relevant to the addressed issue?

In the development of the IBM *Chaoborus* population model, considerable effort has been spent on the implementation of cannibalism as a mechanism of density-regulation that is additional or alternative to intraspecific competition and not shared with most other population models. Indeed, cannibalism and the other implemented mechanisms have been shown to be relevant factors for the dynamics of *Chaoborus* populations. However, various potentially important mechanisms that may limit density-regulation and thus population recovery have so far not been addressed, including environmental and life history effects on the sensitivity of individuals to pesticides (Knillmann et al. 2012a), sublethal effects of pesticides, and interactions with other species (Knillmann et al. 2012b). Given that population recovery seems particularly driven by immigration, further development might be also directed to a more detailed simulation of the recolonization potential.

Are the links between different processes to the variables logical?

Links between the different processes are generally logical, but see comments on the implementation of cannibalism above. As a consequence, survival of life stages other than L1 (subject to density-dependent predation) is generally not affected by any modelled environmental conditions during summer (before dormancy is induced); subpopulations of older larvae are therefore not subjected to den-

sity-dependent "out-thinning". This is probably the reason for the apparently high level of synchronization in the modelled larval development that motivated the authors to switch off intraspecific competition, which rather treats the symptom than the cause.

In the model, the cannibalism rate is determined only by the population density and not by food saturation. This seems in line with observations on other aquatic macroinvertebrates (Koenraadt et al. 2004), and with experiments conducted by the authors showing no shift in prey preference with increasing level of alternative prey (Strauss et al. 2016). However, it seems somewhat inconsistent that cannibalism by L3/L4, but not by L1 larvae has been implemented temperature-dependent. Data for L1 are missing yet and the authors argued that L1/L2 are only present at relevant densities during summer when temperature is close to the test conditions used for parameterization (Strauss et al. 2016). However, this limitation may impede application of the model to scenarios with cooler temperatures in the Northern Zone. Even when the dependency for small larvae is not known, it might have been more logical to assume a similar dependency as for older larvae than to assume no temperature-dependency at all.

Finally, constant food supply over time is assumed in the model. Seasonal variation in food due to the seasonal development of the prey organisms may result in seasonally varying ecological vulnerability of the populations due to times of decreased and increased food stress. This should be particularly considered when the model is applied to consider the effects of different application patterns.

Are the temporal and spatial scales relevant in regard to the problem definition?

The model can simulate ponds of an arbitrary size. It proceeds in time steps of 1 day; this is adequate for the simulation of population dynamics in *C. crystallinus*, but requires specific attention when designing a sub-model for the effects of short (< 1 day) exposure peaks at the individual level. Consequently, the GUTS module applied by Dohmen et al. (2016) was run with hourly time steps.

2.5.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

The model is based on a set of coupled equations that have been thoroughly documented and justified. Apart from cannibalism, many processes in the model are not mechanistic in detail but based on empirical regression, therefore an important assumption is that the patterns observed in the data used for parameterization also hold for the simulated scenarios. This is reasonable as long as the simulated scenarios are close to the (artificial) conditions of the experiments used for parameterization, but results in uncertainty when addressing more complex realistic environments which should be clearly communicated.

The incorporated EC50 toxicity module assigns a random number to each individual at birth that describes the individual pesticide tolerance. As noted by Strauss (2017), this implementation leads to the accumulation of individuals with high tolerance after extended or repeated exposure, i. e. populations develop pesticide resistance. Though resistance has been observed in the field, it is not a worst-case assumption and is typically low in aquatic non-target species (Becker and Liess 2017). A more conservative approach assuming no development of resistance would be to let individuals draw a random tolerance threshold each day of exposure; however, this may require additional information on the variation of effects with exposure profiles (see the case study in section 3.6). This issue does not arise when using the alternative GUTS module to which the population models has been linked.

Are the most important mathematical equations described?

The mathematical equations have been thoroughly described in Strauss et al. (2016) and in Strauss (2017).

Is there a description of the variables and parameters including their meaning and unit?

Variables and parameters together with their units have been comprehensively described in Strauss et al. (2016) and in Strauss (2017).

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

The authors tried to capture the most relevant aspects in the life history of the model species and provided experimental data to close gaps in knowledge, while keeping the model simple enough to be transparent and general. However, decisions on which aspects to model explicitly and in detail and which only implicitly or in much less detail have not been fully justified. E. g., it would be interesting to know why the model development focused on a detailed simulation of cannibalism (which might be simplified to a density-dependent growth rate) and not on other aspects such as reproduction (which could be also simulated as temperature and density-dependent) in order to improve the risk assessment of pesticides.

Are references supporting the equations been provided?

The equations are fully referenced in the supplementary information of Strauss et al. (2016) and in (Strauss 2017).

2.5.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

A flow chart and basic, built-in biological parameter values have been provided in the model descriptions (Strauss et al. 2016, Strauss 2017). A comprehensive description of algorithms and numerical methods is not publicly available.

Is the computer code well readable and is it available?

The computer code is not publicly available but might be obtained from the gaiac institute.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

The modellers did not provide specific information on these items in the publications of the model. However, the model behaviour has been tested with a number of environmental scenarios (Dohmen et al. 2016).

2.5.3.6 The Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

The default environmental scenario presented for model demonstration in Strauss et al. (2016) has been parameterized to a typical Central European pond that fits into the frame of FOCUS exposure scenarios. Applications to other scenarios require a customized environmental parameterization.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

Parameters values in the default environmental scenario such as profiles of temperature and photoperiod, and the migration rate of adults between ponds have been established experimentally or taken from cited literature.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

Pesticide concentrations in the water need to be provided from external exposure modelling. So far, only exposure via the water phase can be modelled. Dietary exposure that may lead to biomagnification in preying larvae cannot be addressed with the model in its current state.

However, the toxicity module has been developed for exposure via the water phase and sediment, not via the uptake of contaminated food, which may be a relevant pathway for bioaccumulating substances.

Is the level of conservatism placed into the scenarios appropriate?

This item cannot be addressed for the model in general. It can be only addressed to a specific application in the context of environmental risk assessment. Scenarios for model demonstration may be optimized for the testing of model behaviour, therefore different criteria may be applied than for the use in risk assessment.

2.5.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

The experiments of the modellers conducted for parameterization have been adequately described and published in a peer-reviewed journal.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Data for parameterization were in most cases obtained from laboratory studies. They therefore captured the performance of the model species under artificial conditions (see section 2.5.3.2), but otherwise appeared of adequate quality.

The presented laboratory experiments on cannibalism rates for parameterization reported no analysis on what actually caused the death of larvae (e. g. based on the appearance of corpses or records of aggressive behaviour). The experiments have been conducted in filtered water without additional food, suggesting that food was limited at all larval densities. Therefore, it could be excluded that the observed increase in larval mortality with density might have been primarily related to competition for food. As a consequence, the observed effect may be indeed related to cannibalism, but also to density stress (crowding or kairomone-related competition for space) as modelled in the IDamP model for *Daphnia magna* (see Preuss et al. 2009a for a discussion). However, the missing differentiation between both death mechanisms does not increase model uncertainty: Indeed, crowding could be more accurately described than cannibalism with the implemented model mechanisms (0 / - relationship, see discussion in 2.5.3.3 above).

Were the estimated parameter values realistic?

Parameter values have been generally drawn from experimental observations (no calibration necessary, except for the conversion from prey density to food saturation) and appear realistic.

Are the data sources sufficiently documented?

All data for parameterization from the literature have been properly cited.

2.5.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

The sensitivity analysis has been adequately documented.

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

A local sensitivity analysis was performed for the basic model without pesticides that is valid only for the tested default environmental scenario. However, scenarios used in risk assessment will probably not differ widely from the selected scenario, and from the results it can be expected that the model is relatively insensitive to changes in the parameters studied. Additionally, individual processes such as migration have been switched on and off to investigate their principal effect on the model output. It can be concluded that the model is quite robust against uncertainties in environmental conditions other than the temperature regime; it is also quite insensitive to sublethal effects of pesticides that alter growth parameters.

However, it would be informative in the model demonstration to assess the sensitivity of the endpoints potentially used for risk assessment, i. e. particularly the recovery time after a given reduction in population size. This would provide a general idea of the uncertainty in risk assessment to be expected when they model may be applied.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

The results have been adequately presented in Strauss et al. (2016); temperature-dependency of the larval growth was identified as the most sensitive parameter.

Has the uncertainty analysis been adequately documented?

No uncertainty analysis has been performed. An uncertainty analysis is most useful for a specific model application and should be performed case-by-case.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

No uncertainty analysis has been performed.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

No uncertainty analysis has been performed.

Uncertainty is propagated to the model results?

The use of random functions for many model parameters covers for the expected uncertainty in input parameters that is provided as probability distributions. This uncertainty is propagated as confidence intervals to the model results. Other sources of uncertainty (e. g. structural) cannot be propagated.

Have confidence intervals been estimated and has this information been used in further model use?

See above.

2.5.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

No performance criteria have been predefined before model validation. The aim was probably to reproduce the observed patterns in population dynamics as close as possible.

Are the model outputs that are compared relevant in view of the problem definition?

In Strauss et al. (2016) the basic model without pesticide effects has been validated to evaluate the structural integrity of the model (pattern-oriented modelling, Topping et al. 2010a). This analysis is useful to assess the structural integrity of the model.

Additionally, the ability of the model to predict appropriate population recovery after pesticide exposure has been validated in Dohmen et al. (2016). The analysis of this endpoint is highly relevant because it will be possibly the most frequently used modelling output for risk assessment.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

The data described in Strauss et al. (2016) were generated from two outdoor microcosm experiments that have been designed for this purpose and met scientific standards. The data in Dohmen et al. (2016) result from an anonymous mesocosm study for the registration of a pesticide. The study is not publicly available but a summary has been provided in the appendix of the publication. Unfortunately, only few data points were available that impedes the comparison of predicted and observed population dynamics in time.

Is the dataset relevant in view of the problem definition?

The basic model without pesticide exposure was tested with a cohort experiment (all individuals of same age) and with populations that were allowed to develop undisturbed in outdoor microcosms (Strauss et al. 2016). This way, temporal patterns of the predicted population dynamics were validated to evaluate the structural integrity of the model. However, the tests were done under comparatively benign environmental conditions (no challenging environmental stress). Under those conditions, the model was quite insensitive to changes in most parameter values (probably because population dynamics were buffered through density-regulation, see sensitivity analysis above). Therefore, it could be informative to validate the model also under other environmental conditions with increased natural stress. Under those conditions the model may become more sensitive to parameter changes because the buffering from density regulation is limited.

Dohmen et al. (2016) first tested the ability of the incorporated GUTS module to predict individual-level effects for exposure profiles that deviate from the constant exposure used for parameterization.

This approach is a very reasonable first step and in accordance with the requirements for GUTS applications in Tier 2C testing (EFSA PPR 2018). However, the actual aim of the study was the validation of the predicted population recovery after acute effects. The mesocosm study used for the validation was conducted in a test facility where exposed ditches are typically closely neighboured to non-exposed control ditches, and no nets were reported to limit migration. It seems that this design may have provided an unnaturally high potential for the recolonization of *C. crystallinus*, which is limited to an active migration distance of < 1 km (Berendonk and Bonsall 2002). Unnaturally high recolonization may have hindered long-term effects after exposure to low concentrations that could have been observed under more realistic field conditions. Therefore, it was mainly demonstrated that the model is not overly conservative, i. e. that it predicts no long-term effects under conditions at which such effects have been neither expected nor observed. Based on the precautionary principle, it would have been more informative to demonstrate that the model is conservative enough, i. e. that it reproduces long-term effects that actually have been experimentally observed.

Is the fit of model output to the data good enough?

In Strauss et al. (2016), the observed patterns in population dynamics are generally reproduced reasonably well, though in microcosms the developmental peak of L2 larvae was predicted ca. 25 d too early. In Dohmen et al. (2016), population densities measured at days 1 – 6 after exposure showed considerable variation in the two replicates per concentration, but GUTS tended to overestimate acute effects at day 1 after exposure but to underestimate more-long-term effects at day 6 e (see Fig. 8). It is unclear whether this was a result from inaccuracy in the fate modelling or in the TKTD modelling, because no measurements of pesticide residues in the mesocosm have been presented. Predicted long-term effects on the population size after 10 – 44 d seem to match the observations well, except that the population model overestimated recolonization after repeated exposure to the highest concentrations at which all larvae were locally extinct for several days after exposure.

Has the performance of the model been reported in an objective and reproducible way?

In Strauss et al. (2016), the validation of the basic model has been presented in an objective and reproducible way. In Dohmen et al. (2016), only few data points are available for the matching of predictions vs. experimental observations. Only the mean observed and predicted population sizes have been compared 10 – 44 d after pesticide exposure, without a measure of variation presented. This complicates a quantitative assessment of the model fit. It would have been also informative to compare the predicted and observed population structure, which is often more sensitive than the overall population size (see validation of the basic model and also Liess and Foit 2010a).

2.5.3.10 **Model Use**

Is a user manual available?

A user manual is not publicly available.

Have all aspects of the modelling cycle been documented?

A full model description according to the ODD protocol is available in the supplementary information of Strauss et al. (2016) and updated in (Strauss 2017). An extended documentation of the model development according to the TRACE framework (Grimm et al. 2014) is currently not publicly available.

Has a summary sheet been provided by the modeller?

No summary sheet as recommended in the Sci. Op. on GMP (EFSA PPR 2014b) has been provided for the general presentation of the model to the public. It may be submitted as part of a dossier.

2.5.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

The model has been published in a scientific paper and is presented at the gaiac website: https://gaiac-eco.de/en/modelling/population-models-ibms/ibm-chaoborus. The modellers are available for dialogue.

Is a version control system implemented?

A version control system has been implemented.

2.5.3.12 Overall Judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

The IBM *Chaoborus* population model has been developed to investigate population effects of pesticides on the predatory freshwater macroinvertebrate *Chaoborus crystallinus*. Specifically, the model is intended to supplement Higher Tier studies with the simulation of effects in populations that experience additional and more realistic environmental scenarios as compared to those in mesocosms. The model species is generally sensitive to insecticides and is a common and ecologically relevant predator in fishless small standing waters Risk assessment based on *C. crystallinus* is, however, not protective for univoltine invertebrates and should be supplemented with additional studies.

Generally, the model seems to be a reasonable compromise between complexity and payoff and has been parameterized with well-documented scientific data. However, some inconsistencies have been identified in the model mechanisms describing cannibalism: Because older larvae do not benefit from preying on 1st instar larvae in the model, indirect effects of pesticides on older larvae via the food chain cannot be captured. However, indirect effects are likely to occur in real populations because 1st instar larvae can represent an important food source for older conspecifics, and because young larvae are often particularly sensitive to insecticides.

Additionally, the selection of implemented mechanisms may be somewhat biased towards processes that potentially lead to an underestimation of the risks of pesticides: In the model, population size seems buffered against acute and sublethal pesticide effects mainly through two mechanisms of density regulation: density-dependent mortality from cannibalism, and food-dependent pace of larval development. Together with the simulated recolonization (and with resistance development in case of repeated exposure), these mechanisms allow populations to quickly recover from pesticide effects under optimum conditions. In contrast, in the current version the model does not capture sublethal and delayed effects which are likely to occur along with acute lethal effects and can significantly decrease the potential for population recovery.

Finally, populations in the field experience additional stressors (e. g. antagonistic species, drought and heat stress); such stressors may considerably lower density-regulation and thus limit the potential for recovery, particularly if they interact with sublethal effects. Such additional (biotic or abiotic) stressors have not been explicitly simulated; they are also not covered implicitly by the parameterization of the biological model part, because the model has been parameterized mainly with laboratory studies

under artificial conditions. Models parameterized with field data aggregate over many of such potential stressors but are limited to comparable ecological conditions. Models parameterized with laboratory data on a number of separate stressors can be used in a more flexible way, but carry the risk that important stressors or their interactions may have been missed.

A clear strength of the IBM *Chaoborus* population model is the comparably extensive work that has been done on model validation. The model reproduced patterns of population dynamics in mesocosms without toxicant exposure well, except for the timing of the 2nd instar development. A local sensitivity analysis showed that under relatively favourable conditions, the model was quite insensitive to changes in most parameters, suggesting that the structural integrity is acceptable. However, it would be informative to test the model also under harsher and more natural conditions where it may become more sensitive. Unlike other population models, the IBM *Chaoborus* population model was also subjected to a validation of predicted population recovery after pesticide effects. Though the comparison was difficult due to the limited available observations, model predictions seemed to match the observed population recovery in mesocosms. Only after exposure to high concentrations that drove local populations to extinction, recolonization was overestimated by the model. In these mesocosms, antagonistic species might have established and impeded recolonization that are not covered by the model. However, the mesocosm study used for validation showed no long-term (> 8 weeks) effects to be reproduced, probably due to an artificially high potential for recolonization in the mesocosms. Therefore, it was rather demonstrated that the model is not overly conservative (no prediction of long-term effects when they have not been observed) than that the model is conservative enough (prediction of long-term effects when they have been observed).

Given that some potential for an underestimation of the overall risk has been identified for the IBM *Chaoborus* population model due the limited ecological processes simulated, the model may be best applied in a probabilistic way: Effects of different risk mitigation measures may be compared, even if the overall risk was underestimated. Additional information on the margin of safety may be obtained when the model is re-parameterized with various worst-case assumptions. E. g., the predicted population effects and recovery might be considered acceptable and not sensitive to increased acute sensitivity and to decreased individual development in the model. In this case, low risk might be concluded also when additional stressors may increase the acute sensitivity and when sublethal effects may delay growth, although the model has not been explicitly designed to cover these mechanisms.

2.5.4 Qualitative Assessment of Uncertainties

2.5.4.1 Potential for Underestimation of Real Risk

- The model has been parameterized for *C. crystallinus* which is generally sensitive to insecticides, but shows low ecological vulnerability due to its comparably high capacity of reproduction and recolonization. Conclusions drawn from modelling *C. crystallinus* may be therefore not protective for sensitive species with long generation times and / or low reproduction rates and / or low migration ability.
- Cannibalism has been implemented only as density-dependent mortality of 1st instar larvae. In the model, preying larvae do not benefit from the prey, thus pesticide effects on 1st instar larvae serving as food source are not indirectly transmitted to preying larvae.
- In its current version, the model does not capture sublethal or delayed effects that may result from a trade-off between resources used for development and for detoxification and from behaviour modification. Capturing such effects may require the replacement of the currently fixed developmental parameters (individual developmental rate, mortality during moulting, number of eggs) with additional modules for individual-level effects that may be fitted using long-term toxicity tests on pulsed (and possibly chronic) exposure.
- The model has been parameterized using data on populations under favourable laboratory conditions. Therefore, environmental (biotic and abiotic) stress that has not been modelled explicitly is also not included implicitly. The only environmental stress explicitly modelled is food limitation and the seasonal effect of day length and low temperature on development and mortality (during dormancy). Additional stressors such as heat stress, drought and antagonistic species may decrease the recovery potential of *C. crystallinus*, particularly when interacting with sublethal effects.
- The built-in modules for individual-level effects may be typically parameterized with data from acute toxicity tests without feeding. Therefore, contact with contaminated water (and probably sediment) is the only exposure route considered, but no uptake with food. Additionally, the model does not account for the potential of biomagnification that may be associated with cannibalism. Therefore, the model should be limited to non-accumulating pesticides.
- The built-in dose-response approach for individual-level effects has been implemented such that the population will develop resistance to repeated exposure, which does not reflect a worst-case assumption.

2.5.4.2 Potential for Overestimation of Real Risk

- The built-in toxicity modules do not consider the bioavailability of pesticides. Bioavailability of pesticides may be lower in the field than in toxicity tests without sediment. Exposure profiles or dose-response curves thus may be adjusted before being used as model input.
- The built-in modules for individual-level effects are typically fitted with data from acute toxicity tests without feeding; in this case they do not account for growth during exposure which may result in the dilution of the internal pesticide concentration. This may result in the overestimation of effects from slowly-acting pesticides. This issue is not relevant when using the alternative TKTD module.

2.5.4.3 Potential for Uncertainty in Either Direction

- Migration rates between exposed and non-exposed populations result from a multitude of landscape properties and are probably highly variable in time and space. Immigration may be therefore estimated only with high uncertainty but has a strong effect on the predicted overall population recovery.

- In the model, only 1st instar larvae are considered subject to cannibalism. No information has been provided on the potential of cannibalism for older larval instars in *C. crystallinus*, though cannibalism on older instars is known from related species such as mosquitoes (El Husseiny et al. 2018). 2nd instar larvae are not considered to prey on 1st instar larvae in the model which should be justified.
- Most parameter values are drawn from probability distributions which enables an uncertainty analysis based on Monte-Carlo simulations. As in all probabilistic models, the generated confidence intervals depend on the assumption that the selected distributions reflect the typically unknown distributions of natural variability.

2.6 MASTEP

Evaluation by Jeremias Becker

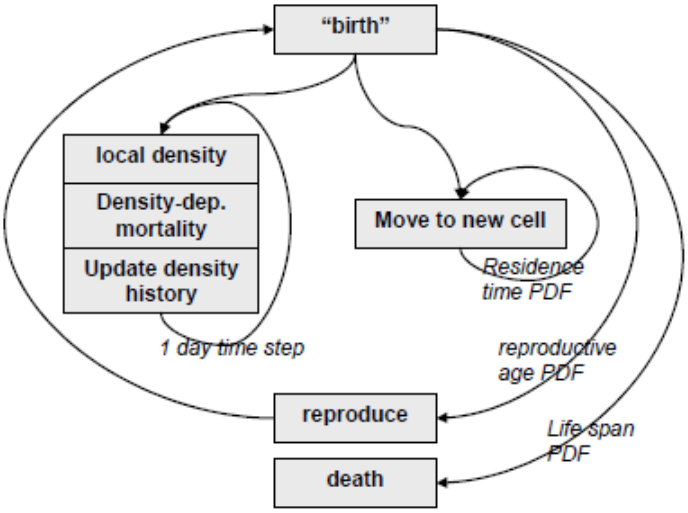
2.6.1 General Information

2.6.1.1 Background and Concept

MASTEP (Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides) is a spatially explicit, individual based population model developed by Van den Brink et al. (2007a)[10]. The aim was to address population effects of pulsed pesticide exposure and the subsequent population recovery in the water louse *Asellus aquaticus*.

The model includes a simplistic life cycle of the model species, with each individual being described by only three properties: identity number, generation number and history of experienced density stress. The simulated life history includes reproduction, mortality and movement. At birth, each individual is randomly assigned to a life span and an age for reproduction drawn from a probability distribution. No life stages are distinguished and each individual reproduces once when it reaches the age of maturity; therefore, only females are simulated. The clutch size decreases with intraspecific competition, quantified as the time an individual has spent on cells that it shared with other individuals during its life. Intraspecific competition also introduces a probability for non-natural death, which was used mainly as a tuning parameter to obtain model populations fluctuating on a desirable density level (Van den Brink et al. 2007a).

Figure 9: MASTEP – Flowchart for the Simulation of Individuals



Overview of life history processes in MASTEP. Boxes show events in life history. Description of arrows in italics show the time delay after which processes take place: residence time, reproductive age and life span are drawn from probability density functions (PDF). Arrows without text point to events that take place 'immediately' (time delay of zero). Individual-level effects of pesticide application was scheduled as a separate event (not shown). Graph reproduced from Van den Brink and Baveco (2009)[11].

[10] Van den Brink, P. J., J. M. Baveco, J. Verboom and F. Heimbach (2007): An individual-based approach to model spatial population dynamics of invertebrates in aquatic ecosystems after pesticide contamination. Environmental Toxicology and Chemistry 26(10): 2226-2236.

[11] Van den Brink, P. J. and J. M. Baveco (2009). MASTEP, an individual based model to predict recovery of aquatic invertebrates following pesticide stress. Book chapter in: Ecological Models for Regulatory Risk Assessments of Pesticides: Developing a strategy for the future. P. Thorbek, V. E. Forbes, F. Heimbach et al. SETAC and CRC Press.

The simulation proceeds in time steps of one day and was typically run for one year. During this time, two main reproduction seasons appear. The individuals randomly walk in a two-dimensional grid with a cell size of 1 m^2. Each cell is attributed to either land or water and can be contaminated with a pesticide concentration predicted by the fate module. Otherwise the environment is homogeneous. The module for random movement was tuned such that on average, an individual moved in one third of its lifetime to a randomly selected neighbouring cell. The model can be run with two principal types of scenarios: standing water and (fast) flowing water. In flowing water, 1 % of all movements are randomly turned into drift events, i. e. the individual jumps to a random cell downstream. The drift distance is drawn from an exponential distribution, by default with an average of 10 m.

When an individual enters a contaminated cell, it is subjected to a probability of immediate non-natural mortality. The mortality risk increases with the concentration following a dose-response curve. MASTEP is intended for fast-acting, non-accumulating pesticides (Van den Brink et al. 2007a, Van den Brink and Baveco 2009). A fate module that predicts the pesticide concentrations on each cell over time was incorporated to demonstrate the operating principle of the model. However, the model can be linked to a separate fate model to address more realistic exposure scenarios.

The model has been parameterized using expert judgement based on an extensive literature review on *A. aquaticus*. If MASTEP is applied in environmental risk assessment (ERA), the authors recommended that the modelling is supported with a mesocosm study to improve the parameterization. Based on a case study, the authors suggested to assess the mortality in the mesocosms 10 days after pulse exposure; these data may be used to directly fit the toxicity module for individual-level effects to observations under semi-natural conditions (Van den Brink et al. 2007a).

2.6.1.2 **Status of the Model**

MASTEP was first published in 2007 (Van den Brink et al. 2007a) with a comprehensive documentation following the ODD protocol (Grimm et al. 2006). In 2009, a shorter documentation followed (Van den Brink and Baveco 2009). In these publications, the model has been applied to the FOCUS standard scenarios for Dutch drainage ditches, ponds and streams mainly to demonstrate the operating principle.

MASTEP was applied in a number of scientific studies to analyse general patterns in the effects of model mechanisms, environmental scenarios, and model species on the predicted population recovery after exposure.

Galic et al. (2012) developed a modified version of the model to address basic research on the effects of structural and spatial heterogeneity of contamination on the recovery potential of a (meta)population. Main differences to the original version include the introduction of body size as a new state variable which depends on local population density and determines the number of offspring, and the definition of specific time frames for reproduction in spring and autumn. The authors concluded that population recovery is mainly driven by the timing of exposure in relation to the reproductive season. Additionally, high connectivity between contaminated and non-contaminated patches decreased recovery times through recolonization, both for the local subpopulations and for the overall modelled population. Local subpopulations in an exposed stretch recovered faster than the overall population, indicating that pesticide effects can indirectly affect also non-contaminated stretches through a disturbed metapopulation dynamic. The authors therefore discussed the need for a more coherent definition of recovery.

Focks et al. (2014b) coupled the modified MASTEP with the CASCADE-TOXSWA fate model and largely extended the size of the landscape (10 km^2 of Dutch ditches) to simulate metapopulation dynamics. The authors provided a detailed documentation of this MASTEP-regional model following the TRACE

protocol (Grimm et al. 2014). Effects in this model were highly sensitive to the assumed individual sensitivity and drift of the pesticide. Focks et al. (2014a) applied MASTEP to scenarios with two pesticides and concluded that the risk of short-term effects added up for both compounds, whereas population recovery was not significantly affected by an additional compound.

Baveco et al. (2014) parameterized MASTEP to four taxa with different traits (*Asellus*, *Gammarus*, Chironomidae, Ephemeroptera) and coupled the population model with a threshold-damage module (a special case of GUTS-SD; MASTEP-TDM). Modelling results for different scenarios were compared to simpler modelling approaches including MASTEP with a classical concentration-effect (= dose-response) module (MASTEP-CE), and a non-spatial logistic growth module coupled to the same concentration-effect module (Logistic-CE). MASTEP-CE and Logistic-CE predicted similar population recovery times when pulse exposure was spatially homogeneous or when individuals were highly mobile (*Gammarus*) so that spatial effects were blurred soon. Local extinction in contaminated patches, in combination with low movement (high local intraspecific competition due to aggregation, and low recolonization), led to longer recovery times in MASTEP-CE as compared to LOGISTIC-CE; note that the non-spatial logistic model averaged over exposed and non-exposed patches. Predictions of MASTEP-CE and MASTEP-TDM on acute population decrease and on recovery times depended highly on the way in which the toxicity modules were implemented and parameterized. In the study, the CE module was fitted to mortality observed after 96 h but executed at a single day of peak exposure, whereas the TDM module was fitted to data from different time points and executed daily for the current exposure concentration. Under these conditions, MASTEP-CE predicted higher acute mortality but faster population recovery than MASTEP-TDM, because in MASTEP-CE no individuals continued to die after the day of peak exposure. The authors concluded that the more spiked the exposure in time becomes, the more the outcome of TDM and CE models coincide.

Galic et al. (2014) also coupled MASTEP with a TDM and a CE module and adapted the model to the amphipod *Gammarus pulex*. The authors concluded that MASTEP-CE provides more conservative predictions on acute effects and recovery than MASTEP-TDM, if parameterized with 96 h or 48 h acute toxicity tests in which the actual exposure lasted for less than 50 % of the test duration (fast-dissipating pesticides). If exposure lasted longer, the CE module may underestimate mortality due to delayed effects that can last longer than the test duration but may be correctly predicted by the TDM module. The results demonstrate the relevance of selecting an appropriate toxicity module in population models.

Finally, Dohmen et al. (2016) applied MASTEP for *Gammarus pulex* incl. GUTS to an anonymized pyrethroid. Predictions were compared to the results of a classical risk assessment approach and to predictions of two other population models (IBM *Chaoborus* population model, IDamP) that were also coupled to GUTS. The authors concluded that the simulated effects were negligible only in the presence of mitigation measures, which coincided with the results from the classical risk assessment approach.

The code and a standalone program of MASTEP is not publicly available, but information can be found on the website http://www.mastep.wur.nl/. In this evaluation we focus on the original MASTEP model for *Asellus aquaticus* (Van den Brink et al. 2007a, Van den Brink and Baveco 2009).

2.6.2 Model Description

2.6.2.1 Problem Definition

Context in which the model will be used

Higher Tier studies for the aquatic risk assessment of pesticides often use mesocosms to refine the risk on freshwater invertebrate communities. These testing facilities are typically exposed more or less homogeneously in space, run for a limited time and not connected to non-exposed refuges from which fully aquatic organisms might recolonize the exposed habitat. The Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides – MASTEP – has been originally published by Van den Brink et al. (2007a) to support Higher Tier studies by the simulation of population effects and recovery through reproduction and recolonization when exposure varies in space and time. The model was developed for the water louse *Asellus aquaticus*; later MASTEP has been adapted also for the amphipod *Gammarus pulex* (Galic et al. 2014) and other freshwater macroinvertebrates Baveco et al. (2014).

Specification of the question(s) that should be answered with the model

MASTEP has been developed to address the magnitude and duration of effects on the population density of the model species at local to landscape scale. Following the European framework for risk assessment, the aim is to assess the maximum decrease in population density and population recovery times in realistic landscapes. This way, specific protection goals for freshwater invertebrates can be addressed that have been defined at the population level.

Specification of necessary model outputs and protection goals

For aquatic invertebrates, the protection goal is usually defined at the population level. either negligible effects (Ecological Threshold Option) or acceptable short- to mid-term effects (Ecological Recovery Option), e. g. effects of medium magnitude that last for up to 8 weeks on an edge-of-field scale (EFSA PPR 2010, EFSA PPR 2013). The required model output is therefore the size and structure of a contaminated and a reference population in time and space.

Domain of applicability of the model

MASTEP is limited to the effects of fast-acting, non-persistent insecticides (Van den Brink et al. 2007a) on *A. aquaticus* and *G. pulex* in a stable freshwater environment. The model may be coupled to FOCUS$_{SW}$ fate modelling for input, but the built-in spatially explicit exposure modelling is beyond the standard FOCUS fate modelling. The default parameterization was done for *A. aquaticus* in Central European ditches. Without modifications in the model design, re-parameterization to other species is possible but limited to invertebrates with a fully aquatic life-cycle. Actually, the model has been also parameterized to flying insects such as Chironomidae and Ephemeroptera (Baveco et al. 2014) , but in these cases the aerial life stages were excluded. The model is intended for a simulation time of up to one year, but may be run longer (the modelled environment is not subject to seasonality).

Why is the model being used?

The purpose of the model is to quantify population effects and recovery after spatiotemporally variable pesticide exposure (Van den Brink et al. 2007a). The specific question to which the model will be applied depends on the uncertainties identified in the risk assessment of a specific substance or product.

What protection goal is being addressed?

The selected protection goal depends on a specific model application and cannot be addressed for the model in general. As outlined above, the typical protection goal to be addressed with the IBM *Chaoborus* population model is no long-term population decline beyond 8 weeks.

What outputs are required?

Addressing the potential protection goals outlined above will require the simulation of the size and structure of a treated and a control population across time (and space, if spatial variation in exposure is considered relevant). Relevant endpoints that can be addressed with the model include the risk of populations becoming extinct and the time for population recovery after exposure.

How was the species chosen?

The authors considered *A. aquaticus* as a conservative model species because it is comparably well-studied, sensitive to pesticides, and ecologically vulnerable due to a low potential for population recovery based on low population growth rates and no aerial migration (Van den Brink et al. 2007a).

Which other species/groups are being covered by the chosen one(s)?

Simulations on *A. aquaticus* (and *G. pulex* in a newer version of MASTEP) potentially cover also other freshwater macroinvertebrates with similar or lower acute sensitivity to pesticides, a fully aquatic life cycle, the ability to drift with the current and a comparable movement behaviour. Due to the simplistic modelling of life history, MASTEP seems not highly species-specific and might be parameterized also to other freshwater macroinvertebrates with comparable life cycles without fundamental changes in the model design. However, many freshwater macroinvertebrates show aggregated spatial distribution patterns due to microhabitat selection, which is not represented in MASTEP.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Galic et al. (2012) applied MASTEP to different scenarios to analyse the behaviour and sensitivity of the model. However, model predictions have never been tested with observations from experiments or field studies.

2.6.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

The authors collected and cited a large number of available studies on the life history of *A. aquaticus* (and *G. pulex*), including both laboratory and field studies. Although some life-cycle characteristics of this species, like age and number of offspring, are known from detailed studies, little information is available for the effects of density stress and walking behaviour (Van den Brink and Baveco 2009).

Assessment of the quality of the data

The cited literature studies have been scientifically peer-reviewed. Few studies have been conducted in the laboratory under artificial conditions, limiting the suitability for parameterization of the model for field populations. The authors provided no specific discussion on the quality of the data.

2.6.2.3 **Conceptual Model**

Description of the model concepts including a diagram

MASTEP is a spatially explicit individual based population model (IBM) with a simple modelled life cycle (Van den Brink et al. 2007a). At the time of hatching, each individual is assigned to an age at reproduction, the minimum date of reproduction in the year, and a lifespan, that are all drawn from probability distributions. Each individual reproduces once when both the age at reproduction and the minimum date of reproduction in the year is reached (only females simulated). In the default settings, this results in two reproduction peaks around the days 120 and 190 of the year. Individuals can die due to the end of the life span, the currently experienced density stress, and due to pesticide exposure. Each individual reproduces once when the age at reproduction is reached (only females simulated). The clutch size increases with age at reproduction and decreases with increasing density stress experienced until reproduction. See the general model description above (section 2.6.1.1) for a flow chart of the model design for individuals.

Individuals randomly walk in a two-dimensional landscape with quadratic grid cells of 1 m^2. Density stress is quantified by the number of individuals that share the same landscape cell. 1% of the random movement steps are drift events in which an individual is transported downstream, with the drift distance drawn from a probability distribution. Individuals can live only in water cells. Water cells differ only in their local population density and in the local pesticide exposure.

The model proceeds in time steps of 1 day. Each time step, individuals exposed to a pesticide are subjected to a probability of pesticide-induced mortality calculated from a dose-response relationship. The model is intended to be used in combination with a mesocosm experiment, from which a dose–response relationship is obtained for the effects of the local pesticide concentration on the local survival. Additionally, the model may be supported with a fate (modelling) study for the distribution of exposure in space and time (Van den Brink et al. 2007a). However, MASTEP provides also a built-in fate module for the dissipation (first-order kinetics) and spatial drift of initial pesticide residues imposed on the model.

Identify the main components and processes in the system

Main components in MASTEP include the individuals, the landscape, and the population (separated in local populations of each cell, subpopulations in initially exposed and initially non-exposed stretches, and the overall population). Main processes include survival, intraspecific competition, reproduction and dispersal (walking and drift).

How the effects of the chemicals are modelled

In the original MASTEP model, the amount, timing and location of exposure (for each grid cell) are part of the user-specific scenario (Van den Brink et al. 2007a). MASTEP can be connected to fate models such as CASCADE-TOXSWA that provide exposure as input for MASTEP (see examples in the general model description above). Additionally, a built-in fate module can calculate initial exposure after a pesticide application for each water cell based on the distance to a treated field cell. In the following simulation steps, the pesticide dissipates using first-order kinetics and additionally moves with the current to neighbouring cells. Therefore, initially non-impacted cells can experience peak exposure later in the simulation.

The only pesticide effect modelled in MASTEP so far is acute mortality. The effect is modelled for each cell separately based on the local pesticide concentration. In the built-in dose-response (= effect-concentration, EC) module, acute mortality is calculated using a steep dose-response relationship. The au-

thors proposed to obtain the EC50 and slope parameter values of this equation from a supporting mesocosm study where population density is measured at the time of peak exposure and 1 week later; the population decline from the first to the second observation may then be used to parameterize the individual-level mortality. This approach is limited to fast dissipating pesticides and considered to integrate potential delayed effects that occur up to 1 week after exposure (Van den Brink et al. 2007a). Alternatively, Galic et al. (2014) and Baveco et al. (2014) fitted the built-in EC module to mortality observed after 96 h in Tier 1 acute toxicity tests. The effect (mortality) can be either executed once at the day of the maximum exposure peak (Baveco et al. 2014, Galic et al. 2014), or scaled and executed daily (e .g. by taking the 7^{th} root of the effect size, and applying this scaled effect during a sequence of 7 days following pesticide application; see examples in the case studies of sections 3.6 and 3.8). Alternatively, MASTEP may be coupled to a TKTD module such as GUTS and its variants (Baveco et al. 2014, Dohmen et al. 2016).

How the components and processes are linked

Background mortality and reproduction (clutch size) are density-dependent. The probability of background mortality increases linearly with the current number of individuals in the same cell; the clutch size decreases with increasing history of experienced density stress (average local density experienced since hatching on a daily base) and increases with age at reproduction. Pesticides affect survival. The individual behaviour (movement, drift, time of reproduction) is not linked to environmental effects.

2.6.2.4 Formal Model

Identification of the model variables

Individuals are described by the state variables identity number, generation number, density history, and probably age. In the revised model of Galic et al. (2012), body size has been added. Populations are described by the local density in each cell, the overall average density in non-treated sections and in treated sections of the water body, and by the total average population density.

Identification of the model parameters

All life-history and movement parameters of *A. aquaticus* are stochastic, using random values drawn from probability distributions. Life-history parameters include the life span, the first day in the year of reproduction, and the age at reproduction. Additional parameters define the density-dependency of survival and clutch size, and the age-dependency of the clutch size. Finally, dispersal-related parameters describe the probability of movement to another cell, the destination of the movement, the probability of downstream drift and the drift distance.

Description of the most important model equations or algorithms

MASTEP is based both on equations (for the biological part of the population model and for the toxicity module) and on simple algorithms (for dispersal). Table 8 summarizes the most important biological equations and the dose-response relationship of the built-in toxicity module. Equations for the built-in fate module are not shown.

Table 8: MASTEP – Important Equations

Definition	Equation	Unit
Probability of survival after exposure	$S = [1 + \exp(-b * (\ln(PEC) - \ln(EC50)))]^{-1}$ b = slope; PEC = local pesticide concentration	%
Density-dependency of clutch size	$\beta = \text{minimum}\{2\beta_n, \beta_n*(A/A_n)\}[\log(D_n + 1)/\log(D + 1)]$ β_n = default clutch size; A = age at reproduction; A_n = Average age at reproduction; D_n = default density; D = average experienced density	Nr. of offspring
Density-dependency of survival	$\mu_{dd} = \mu_1 * N$ μ_1 = steepness of relationship; N = local density	(1 / day)

Data from Van den Brink et al. (2007a, edited).

Dispersal by walking is modelled as a jump from one cell to a randomly selected neighbouring cell after residence time. After each movement a new residence time is drawn from a probability density function; dispersal is takes effect again when the residence time has passed. A given proportion of walking events can be turned into drift events: The individual jumps downstream, the drift distance is drawn from an exponential distribution.

2.6.2.5 Computer Model

Description of the model implementation

The model was implemented in NetLogo using VisualWorks Smalltalk (Van den Brink et al. 2007a).

Checking the computer model for errors, bugs and inconsistencies in the code

A description of code verification was not provided.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

We did not find a description of procedures performed to rigorously test the model behaviour. However, the model was applied to several scenarios for demonstration and research, showing reasonable results.

2.6.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context in which the model is run

The environmental scenario is specific for each model application and cannot be described in general. For model demonstration, Van den Brink et al. (2007a) simulated a 30 x 30 m pond with standing water (with 39 additional cells used to connect inflow to outflow), and a 600 m long and 1 m wide ditch and stream as representative examples. The ditch and stream scenario differed only by flow velocity (10 m/d vs. 250 m/d). The last cell of the ditch and stream was connected to the first one, assuming that the simulated situation is identical to the situation in the upstream sections. The first 100 m of the ditch and stream and the whole pond were subject to an exposure event following application of an insecticide (EC_{50} = 16 µg/L) in neighbouring field cells at day 130 of the year (9th March).

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

The scenario was developed to demonstrate how MASTEP may be applied in a Higher Tier study for risk assessment. The size of the water bodies followed the common FOCUS scenarios and the pesticide properties matched a typical fast-acting, fast-dissipating insecticide (Van den Brink et al. 2007a). Due to the model design, the environmental scenario comprised no further abiotic or biotic parameters.

2.6.2.7 Parameter Estimation

Description of the model parameter estimation

The probability function for residence time after a movement step was developed based on experiment measurements of one-dimensional movement of *A. aquaticus* in a 30 * 100 cm wide enclosure (Englund and Hambäck 2004). Results of this study were extrapolated to the two-dimensional movement pattern in MASTEP using a simulation study (Van den Brink et al. 2007a). The effects of density stress on survival and reproduction, and some parameters required for the random walk were based on expert judgement (educational guess) based on a literature review on the ecology of *A. aquaticus*.

Parameters estimated from the literature — what are the sources and why are these appropriate?

Information on population dynamics and reproduction in streams of UK and Germany was obtained from peer-reviewed publications and a PhD study. A laboratory study indicated that density stress affects reproduction, without quantifying the effect.

Parameters obtained from calibration — how and why this was done?

Density-dependency of survival was tuned to obtain populations that fluctuate around population densities considered realistic (Van den Brink et al. 2007a).

2.6.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

The authors evaluated the sensitivity of population recovery to the parameter for the drifting of *A. aquaticus* by running simulations of the stream scenario described above with and without drift events. Recovery of the subpopulation in the contaminated stretch was highly sensitive to the drifting of organisms, whereas the recovery of the overall modelled population was not (Van den Brink and Baveco 2009). No quantitate sensitivity analysis for the probability of drift or the drift distance were documented.

Focks et al. (2014b) coupled MASTEP with an external fate model and observed that modelled recovery was highly sensitive to the pesticide sensitivity of individuals. Galic et al. (2012) subjected their modified version of MASTEP (addition of body size and specific time frames for reproduction) to a more detailed sensitivity analysis, including the effects of density-dependent mortality and growth. The authors concluded that the model output is relatively robust to changes in parameter values and that the predicted recovery times are most sensitive to changes in the daily movement. Additionally, they concluded that population recovery is mainly driven by the timing of reproductive periods (relative to the timing of pesticide exposure) and decreases with increasing connectivity of contaminated and non-contaminated patches,

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

A specific uncertainty analysis is case-specific, therefore most appropriate once the model has been set up for a specific pesticide and protection goal. As a rule, parameters that showed high impact in the sensitivity analysis on population recovery (see above) have the potential to contribute significantly to the uncertainty of the model. In the described model applications, uncertainty in model predictions is indicated by the 95 % confidence intervals from repeated model runs, but the contribution of different sources of uncertainty to this uncertainty has not been analysed; additional sources of uncertainty (such as structural uncertainty) have not been quantified.

2.6.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

Predictions of MASTEP on population dynamics with or without pesticide exposure have not been extensively validated with independent data from mesocosm or field studies yet. As an exception, the speed of recolonization by *A. aquaticus* in a lake was compared to the simulated recolonization of empty patches (Van den Brink et al. 2007a); the simulated speed (166 m/year for the 90th percentile of the population) was slightly lower than the observed speed (200 m / year). Additionally, in the TRACE document provided by Focks et al. (2014b) the modelled seasonal dynamics of population density in a control scenario were matched to observed abundance in the field from www.limnodata.nl. The field data showed very high variation so that matching was difficult; however, the modelled two distinct population peaks in spring and late summer could not be observed in the field data which suggest rather a constant increase in population density until mid-summer, followed by a constant decrease afterwards.

Additionally, MASTEP output was compared to predictions of other population models and to the outcome of a traditional risk assessment approach without effect modelling which all indicated a low risk for the studied scenario (Baveco et al. 2014, Dohmen et al. 2016). However, no demonstration has been published yet showing that MASTEP predicts high risk for scenarios for which an actual high risk has been identified.

Demonstration that the model output provides an adequate match to data patterns

MASTEP was not validated with independent data.

2.6.2.10 Model Use

Explanation of how the model conforms to the requirements set in the problem definition

The authors considered the model fit for regulatory purposes but acknowledged that validation with independent data is missing and that there is substantial uncertainty in parameters related to the movement behaviour and to the effects of intraspecific competition (Van den Brink et al. 2007a, Focks et al. 2014b).

Description how the model works (user manual)

A user manual is not publicly available, but a detailed model description following the ODD format (Grimm et al. 2006) was published in Van den Brink et al. (2007a), and a more condensed description is available in Van den Brink and Baveco (2009).

Description of the pesticide parameters values used in the model

Pesticide parameters are case-specific and therefore not part of the general model. In the scenarios used for model demonstration in Van den Brink et al. (2007a) and Van den Brink and Baveco (2009), parameter values were used that reflect a fast-acting, fast-dissipating insecticide (see section 2.6.2.6 above).

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed for the model in general, but only for a specific application. As an example, in the first publication of MASTEP, Van den Brink et al. (2007a) demonstrated the simulation of exposure and population recovery in a generic pond, ditch and stream scenario (see section 2.6.2.6 above). Simulations started at day 1 of the year with a low initial population density to reduce computational effort. However, with the onset of breeding in spring, population density quickly increased to an average of 10 individuals / m^2, followed by a larger second population peak in late summer in the control runs. Pesticide application was timed at the height of the first reproduction peak. Exposure to four different initial concentrations were simulated, referring to a similar application rate with four different buffer zones (10, 12.5, 15 and 17.5 m) between the field and the water body. In the pond scenario, the initial pesticide concentration was set such that it decreased exponentially with the distance to the exposed field.

Depending on the initial concentration, the pesticide pulse eliminated ca. 10 – 100 % of the subpopulation in the contaminated patches, and ca. 1 – 50 % of the overall population. In the stream (with drift of organisms and of the pesticide), the subpopulation in the first 100 m contaminated stream stretch quickly recovered to control size within ca. 30 days even after local extinction. However, exposure to high concentrations resulted in a second subpopulation decline as compared to the control in autumn, when all populations decreased after the second reproduction peak. The population size in the overall 600 m stream did not fully recover from high exposure until the end of the simulation in late autumn, because the pesticide drifted downstream and affected also the initially non-exposed stream stretch. Therefore, metapopulation dynamics between the initially contaminated and the initially non-contaminated subpopulation were affected for several months and caused the second subpopulation decline in the first 100 m section.

In the ditch scenario (no drift of organisms and of the pesticide), the local subpopulation in the contaminated 100 m upstream stretch took much longer to recover (ca. 90 d after highest exposure) because recolonization was not supported by the drift of individuals. However, the population size in the overall modelled 600 m stretch was only marginally affected and recovered soon, because the pesticide did not drift to the initially non-contaminated 500 m downstream section, so that most of the population was not directly affected from the pesticide.

In the pond scenario, where exposure decreased with distance from the bank considered to neighbour a treated field, effects on the overall population size were larger than in the ditch but smaller than in the stream. This scenario considered no drift of organisms or of the pesticide. Effects following high exposure decreased strongly after the second reproduction period but remained visible until the end of the simulation. See section 2.6.1.2 in the general information for more examples of model application.

2.6.3 Model Evaluation

2.6.3.1 Problem Definition

The regulatory context in which the model is run

Mesocosms used for aquatic Higher Tier risk assessment typically lack some possibilities of population recovery that may exist in the field. Mesocosm studies do not run long enough to observe autogenic recovery (reproduction) over several generations and do not include connected non-polluted water stretches that may serve as sources for allogenic recovery (recolonization) of species with solely aquatic life stages. Therefore, the authors developed a population model to simulate recovery and thus to supplement conclusions drawn from mesocosm studies. However, it should be noted that recovery of freshwater invertebrates is considered acceptable only if populations recover within up to eight weeks (EFSA PPR 2010, EFSA PPR 2013) that are typically covered by the duration of mesocosm studies. Addressing the potential for longer-term population recovery is therefore of limited relevance in the regulatory context.

The question that has to be answered with the model

MASTEP was specifically designed to assess population recovery after spatiotemporally heterogeneous exposure to a fast-acting, non-accumulating pesticide. The modelling aimed at increasing realism in Higher Tier risk assessment at a landscape scale by the consideration of (meta)population dynamics in a realistic landscape.

The available knowledge and data relevant to the risk assessment question

MASTEP puts a focus on population recovery through recolonization, but information on the actual movement behaviour of freshwater macroinvertebrates is scarce. The model uses a homogeneous environment (other than pesticide) and a random walk algorithm so that a rather homogeneous spatial distribution of individuals is produced However, in real macroinvertebrate populations, individuals are often clustered due to microhabitat preferences at the simulated scale of 1 m^2, resulting in higher actually experienced density stress and non-stochastic movement. It is unclear how a higher level of aggregation may affect the simulated recovery potential.

The second focus of MASTEP is population recovery through reproduction. Populations experiencing high intraspecific competition may readily compensate toxicant-induced mortality through decreased competition-induced mortality and increased reproduction, as simulated in MASTEP. However, external biotic stress (predation, interspecific competition) and abiotic stress may considerably decrease density-regulation and thus the recovery potential (Foit et al. 2012, Knillmann et al. 2012b). Such environmental stressors may culminate in long-term population effects at much lower concentrations than those observed without additional stressors (Liess et al. 2013) but are not explicitly modelled in MASTEP. The authors propose that acute effects in MASTEP are calibrated with a mesocosm study, comparing population densities before and 10 days after pesticide application (Van den Brink et al. 2007a). This way, additional stressors that may increase the sensitivity to acute effects in semi-field conditions compared to laboratory conditions are implicitly covered. Mid-term delayed effects on survival are also covered. However, sublethal effects on growth and reproduction are not considered in MASTEP but may be highly relevant for population recovery and likely to occur at concentrations that are lethal to parts of the population. E. g., pulse exposure even to low concentrations of a fast-acting, non-accumulating insecticide can cause delayed effects on survival and reproduction in later life stages for up to 84 days (Liess and Schulz 1996).

The outputs required to answer these questions including performance criteria for the regulatory model

MASTEP provides population dynamics in time and space from which various estimates for the recovery time can be derived. As demonstrated by the authors, there is a need to agree on a definition of population recovery in the context of risk assessment, because local populations may recover faster than a whole metapopulation, and because disturbed metapopulation dynamics may cause a secondary local population decline. MASTEP simulations may aid the decision on such a definition. Additionally, the model output may be used to assess concentration thresholds that do not exceed a desired recovery time in a specific risk assessment, and the margin of safety for these thresholds when different modelling parameters are varied.

The species to be modelled

While MASTEP simulates a single species, the ultimate protection goal of ERA is the community. The model was developed and initially parameterized for the water louse *Asellus aquaticus*. This species was considered to be relatively sensitive to pesticides and to have a comparably low recovery potential due to a long generation time and the missing ability to fly (Van den Brink and Baveco 2009). However, published SSD curves for various macroinvertebrates and pesticides classify the sensitivity of *A. aquaticus* rather as medium (Roessink et al. 2006, Beketov and Liess 2008a, van Wijngaarden et al. 2010, Sánchez-Bayo et al. 2016, Zhao and Chen 2016). Applications of MASTEP in risk assessment should be therefore supported by additional studies that identified the model species as being most sensitive to the modelled pesticide.

Additionally, the ecological vulnerability of *A. aquaticus* seems not very high because its generation time (2 generations per year in Central Europe) is short compared to univoltine insect species, and because there is no pupal stage like in insects, which can be particularly sensitive to long-term delayed effects (Beketov and Liess 2008b). Therefore, *A. aquaticus* may not represent an ecological worst case for freshwater macroinvertebrates. The same limitations apply to the second model species, *Gammarus pulex*.

However, the life history in MASTEP is sufficiently simple and generic so that the model may be parameterized to various species without the need for changes in the model design. Baveco et al. (2014) applied MASTEP even to aquatic larvae of insects with an aerial life stage. This way, autochthonous recovery of vulnerable species may be addressed, though the simulated recolonization processes will likely not hold for species with aerial life stages.

Requirements for the environmental scenarios to be used in the risk assessment

The model is highly generic in terms of environmental conditions and does not explicitly simulate environmental factors other than pesticide exposure. Therefore, parameters related to life-history traits (life span, timing of reproduction and clutch size) need to be adapted to local conditions when the model is applied to scenarios with climatic and food conditions that differ from those in streams of UK and Central European used for parameterization.

The landscape composition (size, shape and connection of water bodies and agricultural fields) and the spatiotemporal variation in exposure can be specified by the user. These options require the development of agreed landscape scenarios for Higher Tier risk assessment.

2.6.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

The model was not developed based on a specific data set. All parameter values were based on a literature review on *A. aquaticus* covering the last 50 years, and on expert judgement. Sufficient data are available for basic life history traits such as the life span, age at reproduction and natural mortality, and for the initial size structure (in the advanced model version of Galic et al. 2012) . However, information on movement behaviour and density-dependence of growth and mortality is very scarce so that some parameter values had to be estimated with high uncertainty (Van den Brink et al. 2007a, Focks et al. 2014b). This is particularly critical for the movement parameters because recovery time turned out to be particularly sensitive to these parameters.

Has the quality of the data used been considered and documented?

The data used for parameterisation have been referenced in Van den Brink et al. (2007a), and the lack of data on movement behaviour and density-dependency has been discussed.

Have all available data been used? If not, is there a justification why this information has not been used?

The authors based the parameterization on a literature survey; we are not aware of relevant studies that might have been excluded.

2.6.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

For freshwater invertebrates, population effects at the edge-of-fields scale are considered acceptable for days to weeks if long-term decline in biodiversity can be excluded (EFSA PPR 2010, EFSA PPR 2013). MASTEP was designed to assess the time for population recovery of sensitive species and can thus address this specific protection goal.

Are the modelling endpoints relevant to the specific protection goal?

MASTEP can provide various endpoints that may be used to address the specific protection goal, such as population recovery times for different pesticide concentrations, the maximum concentration that does not exceed a given threshold for the recovery time, and the margin of risk that this threshold is exceeded when model parameters are varied.

Is the modelling approach justified?

The model was designed as a spatially explicit IBM to address effects of spatial heterogeneity in exposure and population density on the recovery time. Baveco et al. (2014) showed that MASTEP may perform better than a simpler non-spatial model based on differential equations when the mobility of individuals and thus the potential for recolonization is limited. However, due to the scarce information on movement patterns, a spatially non-explicit approach using simple migration rates between contaminated and non-contaminated populations might have been justified,

A general advantage of IBMs is that they can produce stochastic and demographic effects. Stochasticity is implemented in MASTEP for many parameters, whereas the potential for demographic effects is limited due to the simple life history that does not consider life stages. The decision on a specific toxicity

module for individual-level effects (dose-response vs. TKTD modelling) depends on the specific pesticide properties; pesticides that dissipate and act slowly indicate the requirement of a TKDT module which has been developed for MASTEP.

As a pure population model, MASTEP does not consider the community context in which the model species is embedded. In order to be more balanced in terms of the level of protection, the model may explicitly or implicitly cover relevant ecological mechanisms such as interspecific competition and predation that potentially decrease the recovery potential of populations (see 2.6.3.1).

Is the conceptual model logical?

The conceptual model is straightforward and appears logical.

Are the processes included in the model relevant to the addressed issue?

MASTEP considers density-regulation (through density-dependent survival and reproduction) and recolonization (through walking and drift) which are potentially important processes that facilitate population recovery from acute effects of pesticides. However, the model does not consider a number of important processes that potentially limit population recovery such as sublethal effects of pesticides on growth and reproduction, long-term delayed mortality, differences in sensitivity between life stages, and species interactions (see 2.6.3.1).

In the current state, simulated processes in MASTEP therefore seem somewhat biased towards processes that potentially increase population recovery (recolonization) in contrast to those that decrease recovery. This way, modelling results may overestimate the recovery of real populations in the field, and conclusions for the risk assessment of pesticides should be drawn with care.

Are the links between different processes to the variables logical?

In MASTEP, survival and reproduction is decreased through density stress (history of experienced local population density). However, in the model individuals walk randomly and do not avoid density stress by emigration. This may result in a potential underestimation of recolonization of empty patches, as suggested by the comparison of model predictions with recolonization in lakes after a drought (Van den Brink et al. 2007a). Otherwise links between the simulated processes appear logical.

Are the temporal and spatial scales relevant in regard to the problem definition?

The cell size of the landscape grid is 1 m^2. This is expected to resolve the typical range over which *A. aquaticus* interacts (experiences density stress) and moves. The model proceeds in time steps of 1 d. This seems appropriate for the simulation of population dynamics over several months. However, in the fate module this time step may result in an overestimation of exposure when the dissipation time of a pesticide is very short, so that this module may be run in shorter steps.

2.6.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

The algorithms in MASTEP related to the life cycle are simple and do not seem to require specific justifications. The model design implicitly assumes that the environment is constant in terms of food supply and habitat quality (e. g. water availability), and that species interactions, sublethal effects of pesticides and avoidance behavior or active search for suitable habitat is negligible for the simulation. These assumptions have not been fully justified (see above).

From the model descriptions (Van den Brink et al. 2007a, Van den Brink and Baveco 2009) it becomes not fully clear how pesticide-induced mortality was implemented in the model. The descriptions suggest that during each exposure event, every individual draws a random probability value and dies if it is below the threshold given by the dose-response curve. This approach does not build up population resistance over time. If, in contrast, individuals might draw a random sensitivity at the beginning of life and keep this value, resistant individuals would accumulate after repeated exposure events, which does not reflect a worst-case scenario.

Are the most important mathematical equations described?

All the important mathematical equations have been described.

Is there a description of the variables and parameters including their meaning and unit?

Variables and parameters have been described comprehensively.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

The authors stated that they chose an IBM because it can use available data both at the individual and the population level. A pattern-oriented approach was favoured and only the bare minimum of detail in life history was incorporated in order to keep the model generic and easy to parameterize with various mesocosm experiments (Van den Brink et al. 2007a, Van den Brink and Baveco 2009).

Are references supporting the equations been provided?

The equations are not referenced because most have presumably been built by the authors themselves. Background information is referenced in few cases.

2.6.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

MASTEP was developed in VisualWorks Smalltalk (Cincom Systems, Cincinnati, OH, USA), using the EcoTalk modelling framework. A description of the code is not publicly available.

Is the computer code well readable and is it available?

The computer code is not publicly available.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

No check of the internal consistency or a comparison with a benchmark has been reported. However, the model has been applied multiple times and the realism of the predictions has been evaluated by the authors.

2.6.3.6 **Environmental Scenario**

Is the scenario representative for the risk assessment under consideration?

MASTEP may be and has been applied to different environmental scenarios, therefore this item cannot be answered for the model in general. As an example for the proposed application in ERA, here the fictive ditch scenario in the first publication of MASTEP (Van den Brink et al. 2007a) is discussed that has been applied to demonstrate the operating principle (see section 2.6.2.6).

The scenario considers a single peak exposure during the main reproduction in spring, and subsequent recovery for 1 year. A landscape is modelled in which 1/6 of a representative ditch is not directly exposed from neighbouring fields. While the timing of application may be realistic, in real applications the landscape composition might be too optimistic in areas with intense agriculture. However, no standards for landscape composition have been established yet. A single exposure event has been modelled probably for simplicity in this demonstration, while in real applications MASTEP may be coupled to a more realistic series of exposure events predicted from specific fate models (see case studies in sections 3.6 and 3.8). Peak exposure in ditches and small streams typically result from run-off after heavy rainfall and are thus associated with increased water flow, but in the scenario temporal variation in drift of organisms and pesticides is not simulated.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

MASTEP is simplistic, therefore only few parameters can and need to be adjusted to a specific environmental scenario. The scale and spatial arrangement of the landscape was based on the FOCUS standard scenarios for ponds, ditches, and streams. Pesticide properties were set to typical values for a fast-acting, fast-dissipating insecticide. Drift of pesticides and organisms were parameterized using the scarce literature available. Sensitivity to a pesticide was fitted using fictional dose-response data for demonstration purposes.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

If applied in ERA, MASTEP is considered to be linked to a separate fate model that may address various exposure routes for pesticide input into a water body. MASTEP is explicitly limited to fast-acting, non-accumulating pesticides. Accordingly, the main exposure route for individuals within the water body is considered to be contact with contaminated water, and pesticide effects have been related to water concentrations. However, the option of delivering the complete fate part to an external model provides the flexibility to model also exposure via sediment, if the toxicity module can be fitted to available toxicity data (see an example in the case study of section 3.6). However, exposure via food intake and potential biomagnification cannot be addressed.

Is the level of conservatism placed into the scenarios appropriate?

This item cannot be addressed for the model in general. It can by only addressed to a specific application in the context of environmental risk assessment. Scenarios for model demonstration may be optimized for the testing of model behaviour, therefore different criteria may be applied than for the use in risk assessment.

2.6.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

The basic model parameterization for *A. aquaticus* was mainly based on expert judgement and in few cases on calibration (see section 2.6.2.7). No specific calibration process has been documented, but the data used for expert judgement have been referenced. Parameters specific to the scenarios for model demonstration were fictive.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Sufficient data are available on life history parameters such as life span, age of reproduction and number of offspring for *A. aquaticus* in Central Europe. In contrast, parameters on the movement and density-dependent dynamics had to be estimated from very scarce data.

Were the estimated parameter values realistic?

Parameter values for life history traits appear realistic. For movement and density dependency little information is available, but values do not seem to be extremely off from general expectations.

Are the data sources sufficiently documented?

Data sources from the literature are referenced (Van den Brink et al. 2007a) and the quality of data has been discussed by the authors (Van den Brink et al. 2007a, Focks et al. 2014b).

2.6.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

Sensitivity analyses have been described briefly in Van den Brink et al. (2007a), reporting only the main conclusions. A thorough documentation can be found in the supplementary material of Galic et al. (2012).

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

The sensitivity analyses identified that model settings such as the timing of pesticide application and the landscape composition have major influence on recovery times. The results of the analyses are therefore highly relevant for the application of MASTEP in ERA.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

In the different publications, several parameters and settings have been identified that can have a large influence on predicted recovery (sees section 2.6.2.8). Which parameters are actually most influential will likely depend on the specific environmental scenario and cannot be identified for the model in general.

Has the uncertainty analysis been adequately documented?

No uncertainty analysis has been performed. An uncertainty analysis is most useful for a specific model application and should be performed case-by-case.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

No uncertainty analysis has been performed for the model in general.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

No uncertainty analysis has been performed for the model in general.

Uncertainty is propagated to the model results?

Parameter values for most processes such as lifetime, age at reproduction, probability of movement and distance of drift are randomly drawn from distributions to cover both natural variation between simulated individuals and uncertainty in the input parameters. Thus, parametric uncertainty can be propagated to the model results using confidence intervals.

Have confidence intervals been estimated and has this information been used in further model use?

95 % confidence intervals from Monte-Carlo simulations have been presented for some of the applications for model demonstration (Van den Brink and Baveco 2009).

2.6.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

Predicted population dynamics have not been extensively tested with observed data and no performance criteria have been predefined.

Are the model outputs that are compared relevant in view of the problem definition?

(Van den Brink et al. 2007a) compared the speed of colonization that follows from the random walk algorithm to the speed of recolonization observed along a lake shore. The output is relevant because it addresses a model process that is associated with particularly high uncertainty but has a potentially large impact on the predicted recovery time. Focks et al. (2014b) compared modelled and observed seasonal patterns in population density. This output is relevant for the assessment of the structural integrity of the model. However, predicted recovery times as the most relevant output for use in the risk assessment of pesticides have never been tested with observational data.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

The lake study referenced by Van den Brink et al. (2007a) is a peer-reviewed scientific publication. Field data used by Focks et al. (2014b) were obtained from a public website and the quality has not been discussed.

Is the dataset relevant in view of the problem definition?

The data seem suitable for the purpose of model validation.

Is the fit of model output to the data good enough?

The modelled speed of colonization (166 m per year) was somewhat lower than the observed speed (200 m per year). No details on variation or uncertainty have been provided, therefore the degree of

matching is difficult to assess. However, the values differ by less than 25 %, suggesting that matching was reasonably good. The observed population dynamics show high variation so that matching with modelled data is difficult. The field data suggest that the overall magnitude in the seasonal variation of population density may have been reproduced well in MASTEP; however, the modelled distinct two reproduction peaks were not found in *A. aquaticus* living in ditches of the Netherlands, Variation in the field was much higher than in the model, suggesting that various important processes were not covered in the model.

Has the performance of the model been reported in an objective and reproducible way?

No details have been presented for the matching of the predicted and observed speed of colonization. For the matching of seasonal population density, a detailed graph has been reported.

2.6.3.10 Model Use

Is a user manual available?

A user manual is not publicly available.

Have all aspects of the modelling cycle been documented?

The model has been documented in detail according to the ODD protocol (Grimm et al. 2006). Additionally, a short TRACE document is available for the revised model that includes body size (appendix in Focks et al. 2014b)

Has a summary sheet been provided by the modeller?

No summary sheet is publicly available, but may be submitted as part of a specific dossier for risk assessment. However, a short presentation is available online.

2.6.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

The model is presented on http://www.mastep.wur.nl/documentation.shtml together with a contact form, and the authors are available for contact. Addresses can be found in the various publications on MASTEP.

Is a version control system implemented?

Version numbers have not been reported in the different publications on MASTEP, therefore it is not clear whether a version control system has been implemented.

2.6.3.12 Overall Judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

MASTEP has been developed to supplement mesocosm studies with the simulation of the potential for long-term recovery in freshwater macroinvertebrates through reproduction and recolonization, which is limited in testing facilities. The main model species *A. aquaticus* (and *G. pulex* in later applications)

can be generally considered intermediate in terms of pesticide sensitivity and autogenic recovery potential (generation time). Therefore, if MASTEP is submitted for ERA, more ecologically vulberable or sensitive species should be also considered. The simple life history used in MASTEP may help at this point, because it facilitates to apply the model to various fully aquatic species without changing the model design.

The simplicity in the model design also facilitates the understanding of model processes and the setup of different scenarios. However, the decision to not consider different life stages prevents the analysis of structural (demographic) endpoints which typically recover more slowly than numerical endpoints (such as population size, Liess and Foit 2010a). Additionally, the sensitivity to pesticides can vary between life stages, which cannot be considered in MASTEP in its current state.

As a spatially explicit BM, MASTEP puts a focus on individual movement behaviour that provides the basis for the simulated recolonization. This level of complexity contrasts the simple life history and suffers from some data deficiency. Individuals follow a random walk algorithm and do not avoid pesticide exposure or intraspecific competition, which may result in a potential underestimation of population recovery. On the other hand, landscape conditions are considered spatially homogenous (except for pesticide concentrations) so that individuals do not aggregate in suitable microhabitats as may be the case in field populations. Low levels of aggregation may result in an overestimation of population recovery because on average, individuals will experience lower levels of intraspecific competition (or density stress) than in an aggregated setting. Nevertheless, recolonization patterns simulated in MASTEP appear generally reasonable.

Apart from recolonization, population dynamics in MASTEP are controlled by density-regulation. Other stressors such as desiccation, predation and interspecific competition are not considered but may considerably reduce density-regulation and thus limit the potential for population recovery. Additional stressors may also increase the acute sensitivity of individuals to pesticides, as compared to laboratory conditions. However, unlike other population models, MASTEP was not designed to upscale individual-level effects in the laboratory to field populations. Instead, acute effects in the model are supposed to be parameterized using short-term population decline observed in mesocosms. This way, MASTEP implicitly covers some potential effects of additional stressors in (semi-)natural conditions on the acute sensitivity to pesticides. However, sublethal effects and long-term delayed effects that might considerably reduce reproduction and recolonization are not covered by the model.

Weighing up the mechanisms and aspects that potentially result in an underestimation or overestimation of risk (see also the assessment of uncertainties below) suggests that population recovery predicted with MASTEP may not represent a realistic worst case in the field. Therefore, conclusions for the ERA of pesticides should be drawn with care. The model may be more appropriate to support risk management decisions by comparing effect sizes in different scenarios, rather than to predict absolute effect sizes in risk assessment. Additionally, the model may be used to assess the margin of safety for a given threshold when exposure and environmental scenarios are varied.

2.6.4 Qualitative Assessment of Uncertainties

2.6.4.1 Potential for Underestimation of Real Risk

- Sublethal effects on growth and reproduction, and delayed effects on survival in later life stages are not considered but may reduce the potential for population recovery.
- MASTEP does not differentiate between life stages. Structural recovery of populations takes longer than numeric recovery but cannot be assessed with the model. Additionally, some life stages may be more sensitive than others. E. g., sequential exposure may drive populations to extinction if young individuals are most sensitive and the population structure cannot recover in between exposure events. Such effects are not simulated in MASTEP.
- Species interactions (predation, competition) and abiotic stress may decrease density-regulation and thus the potential of a population to compensate acute pesticide effects due to competitive release. However, species interactions are not specifically considered in the model.
- The landscape is considered homogeneous except of pesticide exposure. Heterogeneity in real habitats may lead to a higher level of spatial aggregation of individuals as compared to the simulations. This may impede recolonization due to barriers of unsuitable habitat patches and increase the experienced intraspecific competition leading to decreased population growth.
- The default model species *A. aquaticus* and *G. pulex* may not cover more sensitive or vulnerable freshwater macroinvertebrates.

2.6.4.2 Potential for Overestimation of Real Risk

- In real populations, avoidance behaviour (induced walking or drift) may reduce exposure, and avoidance of density stress may increase recolonization of empty patches. Movement in the model is purely stochastic.
- For fast-dissipating pesticides with very short DT50 values, simulation of dissipation in daily time steps may result in unrealistically long exposure peaks so that the fate module may require shorter time steps.

2.6.4.3 Potential for Uncertainty in Either Direction

- The environment is modelled temporally homogeneous for the model species, except for pesticide exposure. Food depletion before replenishment through leaf fall in autumn may increase intraspecific competition during the pesticide spraying season in summer and affect population recovery. Additionally, pesticide exposure from run-off events may be associated with increased drift, potentially affecting recolonization.
- Most parameter values are drawn from probability distributions to enable an uncertainty analysis based on Monte-Carlo simulations. The generated confidence intervals depend on the assumption that the selected distributions reflect the typically unknown distributions of natural variability. This assumption must be made in all probabilistic population models.

2.7 SpringSim

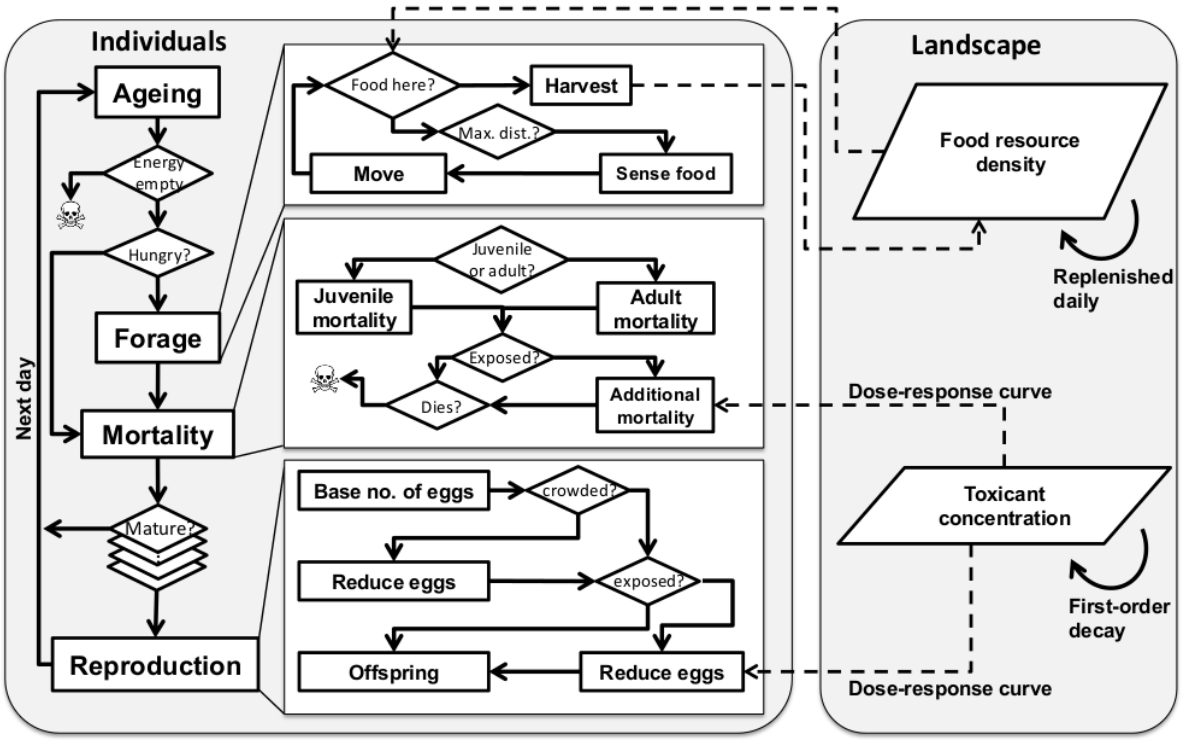
Evaluation by Tjalling Jager

2.7.1 General Information

2.7.1.1 Background and Concept

The SpringSim model, originally developed by Meli et al. (2013)[12], is a spatially-explicit IBM for the soil-dwelling springtail *Folsomia candida*. As with all IBMs, the population part is simply keeping track of all individuals over time. And, as with many IBMs, the life-history traits of the individuals (and the toxic stress on these traits) are based directly on empirical information. The most distinguishing feature of SpringSim is that the individuals sense food level, conspecific density, and toxicant contamination in their vicinity, and base their movement around the grid and their foraging decisions on that (in the most recent version of the model, only food level is sensed by the animals, which is linked to the choice for a coarser spatial grid). Therefore, the model is specifically intended to investigate the influence of spatial heterogeneity, which is especially important in the soil environment as it is less well mixed than water. Furthermore, seasonal changes in temperature and food supply are included.

Figure 10: SpringSim – Flowchart on the Conceptual Model



Conceptual for the individual's schedule of the SpringSim model for the collembolan *Folsomia candida*. Graph from the TRACE documentation of the latest model version SpringSimPy 1.0 (RIFCON 2017)[13].

An overview of the scheduling of the individual's behaviour in relation to the landscape properties is shown in Figure 10. Food is not available in all grid cells, but only in a (randomly selected) fraction of

[12] Meli, M., A. Auclerc, A. Palmqvist, V. E. Forbes and V. Grimm (2013): Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: An individual-based model of springtails. Ecological Modelling 250: 338-351.

[13] RIFCON (2017): TRACE document for SpringSimPy – RIFCON population model for the springtail, *Folsomia candida* (Version 1.0). RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany. Unpublished document.

the cells ('food cells'). For the contamination, various spatial arrangements can be defined (e.g., patches or bands as in Meli et al. 2013).

The life cycle of the animals is simplified to two stages according to age: juvenile and adult female (eggs are only implicitly included; they are not treated as individuals). Due to this simplification, individuals do not have a body size, and do not grow as such; they only mature from juvenile to adult at a certain age. Reproduction starts when the individual has reached the age of maturity, and the animal produces a certain number of broods, with a certain interval between broods, after which it stops reproducing (all temperature dependent). The number of eggs in a brood depends on temperature, 'energy status' of the mother, and local population density. Juvenile survival is represented by a constant probability to survive, and adult survival is implemented through a randomly assigned 'age at death' (both are temperature dependent). In all other respects, juveniles and adults behave the same.

Life history is pieced together from empirical data derived from the literature. A range of traits is considered, such as maturation time, hatching time, eggs per brood, broods per female, time between broods, etc. For each of these traits, every individual selects a set of parameter values from (independent) random distributions at the beginning of each season (original model) or each simulation day (latest version of the model); the distributions depend on the actual temperature. A crude form of 'energy budget' is included, but there is no conservation of mass and energy. Energy is tracked as abstract 'days-to-death'. Feeding increases the energy status: when an individual finds a patch with food, it eats until the food is gone or its maximum reserve capacity is achieved. The energy status is decreased by costs for maintenance and movement, but not by maturation or reproduction (egg production thus does not cost energy in the model, though the number of eggs in a brood is scaled by the mother's energy status). Energy gain by feeding and energetic costs of maintenance and movement are taken independent of body size (i.e., stage) and temperature. This is clearly unrealistic, but as in the real world both the energy sources (feeding rate) and the energy sinks will increase with increasing size and temperature, the net result may not be completely unrealistic. Egg production rate is linked to energy status and local density (crowding) according to independent exponential functions. However, the energy status does not influence growth (i.e., maturation time) or the total number of broods. The energy rules were not based on experimental data at the individual level, but calibrated on two population 'patterns' (only those parameters were calibrated that turned out to be sensitive at the population level). Individuals can move around their world, and make decisions (stochastically) based on sensing food level, and optionally toxicants, in their immediate neighbourhood. Movement is not affected by temperature, but only by energy status, and the food availability in the neighbouring cells. As pointed out above, in the original version of the model, the animals also sensed the density of conspecifics in neighbouring cells, and avoided contaminated patches to some extent. Food is available in some parts of the environment; a small percentage of the area is randomly assigned as 'food cells', and these are restocked at the beginning of the next day (in the last version of the model, restocking probability depends on a seasonal soil-moisture pattern).

Toxicity is included through static dose-response curves for survival, fecundity and hatching success from standard 28-day tests. Linear regressions of endpoints against log-concentration have been used in the original publications, but the latest version uses the more standard log-logistic curves. As the animals are followed on a daily basis, and as they move on and off contaminated patches, this needs to be translated to a smaller time scale. In the original model, this was done by a form of time-weighted averaging. Every hour that an individual spends on a contaminated patch, a toxicity counter is increased, which is used to scale the toxic response observed in the 28-day test. If the counter exceeds 28 days, it remains at the effect level observed in the 28-day toxicity test. There is no way to decrease this counter, so no elimination or repair of damage is considered. In the last version, the model uses an alternative assumption of immediate effects at the level of the 28-day test. As long as the individual stays in a contaminated patch, it will experience the total effect following from the 28-day dose-response curve. When it moves to an uncontaminated patch, the effect is zero again. This is expected to

produce a worst-case estimate for toxicity, although this will not always be the case (instant recovery is unrealistic, and could underestimate toxicity in heterogeneously contaminated environments). No toxicity on time to maturation or total number of broods is included, which would also be impossible to derive from the standard toxicity data with the species. Furthermore, toxicity is assumed to act independently from temperature, life stage, crowding or energy status. The distribution of contamination in the environment can be taken either homogeneous or heterogeneous. For heterogeneous contamination, part of the environment was contaminated with a particular concentration, and another part with a lower or zero concentration. Within the contaminated and uncontaminated area, the concentration was thus taken homogeneous. However, more complex patterns can likely be inserted.

There is a large number of descriptive rules in this model, and the exact rules are not always easily reconstructed from the papers and early TRACE documents. However, the TRACE document for the latest version (SpringSimPy 1.0) is more complete and more structured.

2.7.1.2 **Status of the Model**

The model is programmed in NetLogo (https://ccl.northwestern.edu/netlogo/) and has already sporadically been submitted as part of regulatory dossiers for PPPs. Recently, an analysis with the model, submitted in support of a product registration, was accepted by authorities (Mattia Meli, pers. comm.). The model is currently owned, maintained, further developed, and applied by RIFCON consultancy. It is not publicly available, but it is available (with a TRACE documentation) for authorities on request. For the first version, the NetLogo code was available as supporting information to the publication (Meli et al. 2013). More recently, the model was re-implemented in Python at RIFCON (SpringSimPy), but without a user interface. RIFCON is currently working on further refinement and improvements of the model (RIFCON, pers. comm.).

The first version, which was used for the publications in the scientific literature, focussed on effects of heavy-metal contamination (copper) on *F. candida*. More recently, the model has been modified, and is now presented as specifically developed for risk assessment of pesticides. The update included aggregation to a coarser temporal scale (daily basis instead of considering foraging on hourly basis) and coarser spatial scale (grid cells 10 cm^2 instead of 1 cm^2). Individual parameters are now dependent on the daily temperature rather than on season, individuals now sense only food in neighbouring cells and not conspecific density anymore (which are assumed to be less relevant for the coarser spatial scale). Furthermore, the model now includes a different option for the toxicity calculations: The original model used a crude form of time-weighted averaging. The latest version uses a dose-response curve that is based on data from 28-day standard reproduction studies, and relates the nominal dose to the total effect observed over 28 days. On a daily basis, the actual exposure concentration determines the extent of the response in the model using the total-effect dose-response relationship. This assumption will overestimate the toxicity on exposure to contamination, but will also overestimate recovery (instant recovery when an individual moves to a clean patch of the environment).

The user community of the model seems to be small, and the four published papers seem to be mainly based on the work of a single scientist (Mattia Meli, working at that time at Roskilde University, Denmark).

2.7.2 Model Description

2.7.2.1 Problem Definition

Context in which the model will be used

SpringSim is a model for the population dynamics of the springtail *Folsomia candida*. The original model (Meli et al. 2013) was developed and parameterised for heavy metals (copper), but the last version (as maintained by RIFCON) is geared more specifically towards PPPs. The model would most likely be used to extrapolate from standard toxicity tests with *F. candida* to potential impacts for a population under field conditions. As the authors explain (Meli et al. 2013): "The purpose of the model is to simulate *F. candida* population dynamics and to investigate how they are affected by spatial distribution of toxic contamination in soil, with a special focus on interactions with food availability and local population density."

Specification of the question(s) that should be answered with the model

Questions that deal with the impact of soil contamination on the population of springtails. The model deals with a single population, living in a closed two-dimensional area. The model is spatially explicit, and individuals move around the landscape in search of food. Food and contamination can be defined to be heterogeneous in the landscape.

Specification of necessary model outputs and protection goals

Different types of output can be derived from the model such as population growth rates, spatial distribution, equilibrium population size, or recovery of the population. The paper of Reed et al. (2016) shows how this model could potentially be applied in a risk assessment, for a hypothetical PPP.

Domain of applicability of the model

The model focuses on *F. candida*, and does not consider interactions with other species (e.g., as predators or competitors).

Why is the model being used?

To extrapolate from standard toxicity tests with *F. candida* to potential impacts for populations under field conditions.

What protection goal is being addressed?

Protection goals dealing with population dynamics of *F. candida*.

What outputs are required?

Different types of output can be derived from the model such as population growth rates, spatial distribution, equilibrium population size, or recovery of the population.

How was the species chosen?

F. candida is a common soil arthropod, it plays an important role in the soil ecosystem, it is used extensively as model soil arthropod, is sensitive to toxicants, and it is a standard test organism in ecotoxicology and ERA.

Which other species/groups are being covered by the chosen one(s)?

The model is specific for the springtail *F. candida*. *F. candida* is a parthenogenetic and opportunistic species, capable of high population growth rates under optimal environmental conditions. It may therefore be less representative for sexually reproducing species with a more specialist lifestyle. However, the model could be adapted to other springtail species, if they are similar enough in life history to *F. candida*, or when sufficient information on the life history and behaviour is available.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

See next section.

2.7.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

For the supporting data used in the model, three categories of data need to be distinguished. There are data used to derive the relationships that govern the basic life history of *F. candida* (model parameterisation), such as the age at maturity and brood size depending on the temperature. These data can be considered part of the model as they would remain unchanged for most applications. Additional data would be needed to include the effects of a toxicant on the life-history traits; in the first version of the model, data for copper sulphate were used for this purpose (standard 28-day toxicity data plus egg viability). For the substance-specific data, the data are obviously case specific, but it should be noted that the information content in the standard *Folsomia* test is very low as effects are only scored at the end of the test (and exposure is often only reported as nominal soil concentrations). Finally, data were used for calibration and validation (at the individual and population level).

The data used for calibration (pattern-oriented modelling) comprises two studies in a laboratory setting. One study reports population abundance over time (125 days) at three food levels, and the other study only population growth rate and equilibrium population size. The data used for validation (as in output corroboration) are quite limited: the mean number of generations over one year observed in a greenhouse, population density in a forest at four points in a year, and population growth rates under homogeneous copper exposure, very crudely estimated from a standard 28-day laboratory test (only results at the end of the test used). For the last version of the model (SpringSimPy 1.0), additional data were used for model testing: number of eggs produced per individual over its entire life span, from several studies. Furthermore, the comparisons for two of the data sets was treated as 'verification', and the other two as more strict 'corroboration' (several studies were known during the model development and were implicitly taken into account in the model design). The last version of the model produces somewhat different predictions than the first version in these comparisons to data. This discrepancy is not explained, but likely relates to a number of changes that have been made to the model since then.

Assessment of the quality of the data

For all the types of data used, the quality of the data was not discussed, but references are provided to the literature sources used.

2.7.2.3 Conceptual Model

Description of the model concepts including a diagram

A flow chart of the model was already presented in section 2.7.1.1. The model deals with a single population, living in a closed two-dimensional area (only the top-soil layer of 0-5 cm is considered, which is treated as homogenous). The environment is heterogeneous with respect to food (a certain percentage of grid cells is defined as food cells) and toxicant contamination. The model is spatially explicit, and individuals move around the landscape in search of food. Food dynamics is rather artificial: only some cells contain food, and they are restocked the next day when empty (in the last model version, restocking depends on the season). Factors that influence life history of individuals are food, temperature, crowding and toxicant stress. There is no predation, no competitors (other than the conspecifics), and no diseases.

Identify the main components and processes in the system

The life-history of individuals is captured by a large set of empirical relations (based on experimental studies).

How the effects of the chemicals are modelled

The model is meant to be applied using the data from the standard reproduction test on *F. candida*. These studies can be used to obtain dose-response curves for survival and reproduction at the end of the test duration. In the original model, some form of time-weighted averaging was used, but in the last version, the effect from the dose-response curve is applied to the individuals instantly on a daily basis, based on the actual concentration in the cell that they settle in on a day. There are thus no TKTD consideration in the model.

How the components and processes are linked

See diagram in section 2.7.1.1. The IBM engine subsequently follows all individuals as they develop and move over time.

2.7.2.4 Formal Model

Identification of the model variables

The TRACE document for SpringSimPy 1.0 summarises the state variables as follows. For a landscape cell: location (x- and y-coordinates), food cell (yes/no), food resources (energy units; 0 if non-food cell, 0 to maximum if food cell), treatment status (yes/no), actual substance concentration in soil (mg/kg). For an individual: location (x- and y-coordinates), age (days), developmental status (juvenile/adult), energy reserves (energy units; 0 to maximum), age at previous oviposition, sizes of previously deposited egg batches, number of previous oviposition events, actual exposure to substance (mg/kg).

Identification of the model parameters

The model parameters and their value (incl. sources) were summarised by the authors in their Table 1 (Meli et al. 2013).

Description of the most important model equations or algorithms

The model applies a range of empirical relationships and decision rules to capture the life history and behaviour of the individuals. The IBM engine subsequently follows all individuals over time.

2.7.2.5 Computer Model

Description of the model implementation

The original version of the model was programmed in NetLogo (https://ccl.northwestern.edu/netlogo/), which is a freely-downloadable platform for developing and running IBMs. The model is currently owned, maintained, further developed, and applied by RIFCON consultancy. It was re-implemented in Python at RIFCON (SpringSimPy), but without a user interface. RIFCON is currently working on further refinement and improvements of the model.

Checking the computer model for errors, bugs and inconsistencies in the code

The TRACE documents also specify the steps taken to verify correct implementation (e.g., review of the code, following individuals and comparing their development to calculations by hand, extreme parameters). Details of these checks have not been provided.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

See above. Two quantitative verifications are provided in the TRACE document (RIFCON 2017), sections 5.2 and 5.3.

2.7.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context **in which the model is run**

Different scenarios can be run for the model, in terms of environmental properties and in terms of soil contamination, spatial heterogeneity, etc.

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

This item cannot be answered; a wide range of different scenarios can be run. Biotic/abiotic factors that are included into the model are temperature, food availability, conspecific density, and toxicant stress. Agro-environmental parameters will translate into level and distribution of the toxicant, and possibly food availability. The model only deals with one species, and interactions with other species therefore cannot be simulated.

2.7.2.7 Parameter Estimation

Description of the model parameter estimation

Model parameters for the life history of the individual (and the toxic effects) are based on empirical data. Most parameters have a distribution and temperature dependence that is also based on empirical data. The parameters governing the dose-response curve for the toxicant effect are the same for all individuals and do not depend on the temperature.

Parameters estimated from the literature — what are the sources and why are these appropriate?

Various sources for the data were consulted; references are provided.

Parameters obtained from calibration — how and why this was done?

For several parameters, no relevant literature data could be found (mainly the parameters of the 'energy budget'). Several parameters were subjected to calibration (pattern-oriented modelling), namely

those that were sensitive in relation to the population response; where the response is taken as final population size and average weekly population growth rate.

2.7.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

A limited sensitivity analysis was used for the parameters for which no relevant literature data could be found (mainly the parameters of the 'energy budget'). The parameters that were sensitive in relation to the population response (final population size and average weekly population growth rate) were subjected to calibration (pattern-oriented modelling). The sensitivity analysis was described in detail in the TRACE document of the first model version (supporting information of Meli et al. 2014a).

In Meli et al. (2014b), the authors also performed a kind of structural sensitivity analysis, by comparing their model to a simpler matrix model, and by exploring the influence of heterogeneity in the contamination.

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

No uncertainty analysis is applied. However, the model is stochastic as most parameters have a random distribution; at the beginning of each day (latest model version) or each season (original model version), every individual draws a new set of parameter values from (independent) random distributions, based on the actual temperature. This variation is propagated to uncertainty in the population response, which can be summarised as intervals around the mean model output (see figure in section 2.7.1.1).

2.7.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

Several data sets were used for validation: the mean number of generations over one year observed in a greenhouse, population density in a forest at four points in a year, and population growth rates under homogeneous copper exposure, crudely estimated from a standard 28-day laboratory test (only results at the end of the test used). For the last version of the model (SpringSimPy 1.0), additional data were used for model testing: number of eggs produced per individual over its entire life span, from several studies. Only one of the validation studies includes toxicant stress, albeit perhaps a less relevant one from the PPP perspective (copper, constant exposure), and only in a rather short-term laboratory setting (a standard 28-day test).

Demonstration that the model output provides an adequate match to data patterns

The correspondence of the model to these data sets was fair, although the data basis for these comparisons is very limited.

2.7.2.10 Model Use

Explanation of how the model conforms to the requirements set in the problem definition

This item cannot be addressed in general. The model follows a simple principle whereby all individuals are followed (IBM), and the life history of each individual is based on empirical data for the species.

Description how the model works (user manual).

The model has been described in a range of papers and in several TRACE/ODD documents.

Description of the pesticide parameters values used in the model

This item cannot be addressed in general. Pesticide effects are included as static dose-response curves for survival and reproduction, which is thus case specific.

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed in general.

2.7.3 Model Evaluation

2.7.3.1 **Problem Definition**

The regulatory context in which the model is run

SpringSim is a model for the population dynamics of the springtail *Folsomia candida*. The original model (Meli et al. 2013) was developed and parameterised for heavy metals (copper), but the last version (as maintained by RIFCON) is geared more specifically towards PPPs. The model would most likely be used to extrapolate from standard toxicity tests with *F. candida* to potential impacts for a population under field conditions. As the authors explain (Meli et al. 2013): "The purpose of the model is to simulate *F. candida* population dynamics and to investigate how they are affected by spatial distribution of toxic contamination in soil, with a special focus on interactions with food availability and local population density."

The question that has to be answered with the model

The model deals with a single population, living in a closed two-dimensional area (only the top-soil layer of 0-5 cm is considered). The environment is heterogeneous with respect to food (a certain percentage of grid cells is defined as food cells) and toxicant contamination. The model is spatially explicit, and individuals move around the landscape in search of food. Food dynamics is rather artificial: only some cells contain food, and they are restocked the next day when empty (in the last model version, restocking depends on the season). Factors that influence life history of individuals are food, temperature, crowding and toxicant stress. There is no predation, no competitors (other than the conspecifics), and no diseases.

The available knowledge and data relevant to the risk assessment question

This item cannot be answered for the model in general.

The outputs required to answer these questions including performance criteria for the regulatory model

This item cannot be answered for the model in general. Different types of output can be derived from the model such as population growth rates, spatial distribution, equilibrium population size, or recovery of the population. The paper of Reed et al. (2016) shows how this model could potentially be applied in a risk assessment, for a hypothetical PPP.

The species to be modelled

The model is specific for the springtail *F. candida*. However, the model could be adapted to other springtail species, if they are similar enough in life history to *F. candida*, or when sufficient information on the life history and behaviour is available.

Requirements for the environmental scenarios to be used in the risk assessment

Different scenarios can be run, which are specified by a temperature profile, food level (and its spatial distribution and re-stocking schedule), and contaminant exposure pattern (both temporal and spatial).

2.7.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

Yes, although the data used at the individual level (for parameterisation) and the population level (for testing) is rather limited.

Has the quality of the data used been considered and documented?

For all the types of data used, the quality of the data was not discussed, but references are provided to the literature sources used.

Have all available data been used? If not, is there a justification why this information has not been used?

The data used are relevant, but more data is available that could have been used to strengthen the model. For example, the results of controlled microcosm population experiments (Noël et al. 2006) could be useful for this purpose. The model has not been compared to detailed data at the individual level, i.e., individual traits over time as function of food, temperature and toxicant concentration. Such data sets exist (e.g., Jager et al. 2004, Hamda 2013), and could be used to more thoroughly test the individual part of the model.

2.7.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

This item cannot be answered for the model in general.

Are the modelling endpoints relevant to the specific protection goal?

This item cannot be answered for the model in general. Various endpoints can be obtained from the model.

Is the modelling approach justified?

The general logic of IBMs is to calculate population dynamics by following all individuals over their life cycle. This is obviously very defensible; the main disadvantage will be that model simulations will be very calculation intensive. However, the realism of an IBM will (like any population model) depend heavily on the representation of the individuals and the environment. Firstly, it needs to be stressed that the model only considers a single species, *F. candida*, and thus no interactions with other species. Secondly, the representation of the individuals in SpringSim is descriptive and rather crude. This may be justifiable as long as long extrapolation do not range far from the conditions in the experiments used to calibrate the model; uncertainty in the model output will increase with increasing extrapolation distance.

Is the conceptual model logical?

The modelling logic is to derive the basic individual traits from experimental data, as function of temperature and local density. A crude form of 'energy budgeting' was included, which was tuned to reproduce certain patterns at the population level. The approach taken is logical, but seems rather *ad hoc* and data driven. Furthermore, some logical consequences were not included, such as the effect of temperature and life stage on the energy budget (feeding, maintenance costs, etc.) or the effect of food and toxicants on maturation time. Therefore, it is unclear whether the model is capable of providing a

reasonable representation of an individual's life history in response to food, temperature and density/toxicant stress. Furthermore, all factors are assumed to act independently on the individual, which might introduce additional uncertainty or bias.

Are the processes included in the model relevant to the addressed issue?

Yes, but see remarks on the modelling approach and conceptual model above.

Are the links between different processes to the variables logical?

Yes, but see remarks on the modelling approach and conceptual model above.

Are the temporal and spatial scales relevant in regard to the problem definition?

Yes, although it will depend on the specific problem that the model is applied to.

A strength of the model is that it incorporates spatial heterogeneity (only in the horizontal plane, not in soil depth) and accounts for changes in the environmental scenario (temperature and food availability) over the season. In the original model, the spatial scale consists of 1 cm^2 grid cells, the foraging module acts in hourly time steps, whereas the other individual-level processes are followed on a daily basis. Food is reset on a daily basis (with a certain probability), and there are six seasons in a year (at the start of a new season, new parameters for the individual are selected from temperature-specific random distributions). In the last version of the model, all individual behaviour (and temperature) is on a daily basis, and the temporal scale consists of 10 cm^2 grid cells. There is no limit to the area or time period that can be modelled, but the original model considered a 1 m^2 environment and was run over a year.

The original publication (Meli et al. 2013) included some simple distribution of patches with contamination, but in a subsequent publication (Meli et al. 2014a), more realistic fractal distributions were used. In all cases, contamination in each cell was either present or absent (or 'low'). Food cells are assigned randomly to a certain percentage of the cells in the simulated landscape, and, in the last version, restocked with a certain probability and to a certain level depending on the time of year (to mimic fungal growth as a function of soil moisture) every day (in the original model, food was restocked every day to the maximum level). This type of distribution of food and contamination does not seem to be very realistic, although more reasonable scenarios could be developed for this model.

2.7.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

The model for the individual rests on empirical trait values and regressions, largely taken from the literature. References are provided. Some parameter values were tuned to provide a reasonable prediction for some observed patterns. The most important assumption is thus that the literature data are representative, and that all environmental factors act independently on the individual. The TRACE document for the last version also includes a motivation for several of the simplifying assumptions (e.g., the implicit egg stage and the lack of vertical migration).

Are the most important mathematical equations described?

Yes, there are several ODD and TRACE descriptions of the model, which include the equations used, parameters, units etc., including a justification for the approach taken. The last version (for SpringSimPy 1.0) is clear and complete.

Is there a description of the variables and parameters including their meaning and unit?

Yes, see above.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

An interesting justification of model complexity was provided in the form of a comparison to a simpler model, namely a standardised meta-population matrix model (Meli et al. 2014b). For scenarios with homogeneous contamination of the landscape, both models provided rather similar results. However, for heterogeneous contamination, there were more profound differences, which could be explained from the lack of individual behavioural responses in the matrix model.

Are references supporting the equations been provided?

Yes, see above.

2.7.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

For the first model version, NetLogo Code is provided in the various papers. However, the latest version of the model is not publicly available but can be obtained on request by authorities (it is currently owned by RIFCON), and therefore the code and documentation are not publicly available (a TRACE document for SpringSimPy 1.0 was sent to us on request). Descriptions of the first version are provided following the ODD protocol (Meli et al. 2013) as well as a TRACE documentation (supporting information of Meli et al. 2014a, Meli et al. 2014b).

Is the computer code well readable and is it available?

See above.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

The TRACE documents also specify the steps taken to verify correct implementation (e.g., review of the code, following individuals and comparing their development to calculations by hand, extreme parameters). Details of these checks have not been provided.

2.7.3.6 The Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

That depends on the 'risk assessment under consideration'; the model allows for simulating many different scenarios.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

See above.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

See above.

Is the level of conservatism placed into the scenarios appropriate?

The model itself is not particularly conservative or non-conservative, it is the scenario, the type of output used, and the assessment factors selected that will eventually determine the level of conservatism. Different scenarios can be simulated with the model, in terms of temperature profile, heterogeneity of contamination, food availability, degree of contaminant avoidance (by default excluded in the last version), etc. Some worst-case elements are included into the model, such as restriction of the animals to the top 5 cm of the soil column.

The toxicity module in the last version applies a static dose-response for the toxicant effect: instant effect at the level of the dose-response curve derived from the results of a 28-day toxicity test, for the actual exposure level that the individuals encounters in an environmental cell. The total effect over the 28-day test observation period is applied on a daily basis, and as long as residues exist in the cell where the individual settles for the day. Whether this represents a worst case depends on the exposure scenario. When the concentration is homogeneously distributed in the environment, and rather constant, this will be a worst case (unless the toxic effect is really slow and the EC*x* continues to decrease even after 28 days of constant exposure). When contamination is heterogeneously distributed, or the chemical concentration decreases over time, it is important to note that the use of a static dose-response will lead to instant recovery of the individual. There is no accumulation of the toxicant over time, and no delayed or irreversible effects. It is impossible to decide a priori whether the static dose-response will over- or underestimate the effects, as it will depend on the details of the exposure pattern for the individuals, and on the TKTD properties of the chemical in this species.

2.7.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

The model parameterisation is largely based on literature data for *F. candida*. Data sources are referenced and additional calculations are explained. Several parameters (those relating to the 'energy budget') were tuned to obtain reasonable correspondence to observed patterns.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Yes.

Were the estimated parameter values realistic?

Parameterisation was done for all factors independently (temperature, food, density, toxicants). It is unclear whether this leads to reasonable results for the combination of those factors (as they might interact). Several factors might be important but have not been considered (e.g., effects of food on maturation times, or temperature on the 'energy budget' and movement).

Are the data sources sufficiently documented?

Yes.

2.7.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

A limited sensitivity analysis was used for the parameters for which no relevant literature data could be found (mainly the parameters of the 'energy budget'). The parameters that were sensitive in relation to the population response (final population size and average weekly population growth rate) were subjected to calibration (pattern-oriented modelling). The sensitivity analysis was well described in the TRACE document of the first model version (supporting information of Meli et al. 2014a).

In Meli et al. (2014b), the authors perform a kind of structural sensitivity analysis, by comparing their model to a simpler matrix model, and by exploring the influence of heterogeneity in the contamination.

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

Yes, although it depends on the specific problem that the model is applied to.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

Yes; the most sensitive parameters were subsequently submitted to pattern-oriented modelling.

Has the uncertainty analysis been adequately documented?

No uncertainty analysis is applied. However, the model is stochastic as most parameters have a random distribution; at the beginning of each day (latest model version) or each season (original model), every individual draws a new set of parameter values from (independent) random distributions, based on the actual temperature. This is a rather strange situation, where individuals will change their life-history characteristics from one day to the next. The distributions selected are provided with a reference on which they are based.

The inter-individual variation is propagated to the population response, and summarised as intervals on the mean response.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

See above.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

See above.

Uncertainty is propagated to the model results?

See above.

Have confidence intervals been estimated and has this information been used in further model use?

The model output is given with intervals, representing the range of the model runs. Each run is different due to stochasticity at the individual level.

2.7.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

Some validation was performed but no quality criteria were defined.

Are the model outputs that are compared relevant in view of the problem definition?

The data used for validation are relevant but rather limited: the mean number of generations over one year observed in a greenhouse, population density in a forest at four points in a year, and population growth rates under homogeneous copper exposure, crudely estimated from a standard 28-day laboratory test (only results at the end of the test used). For the last version of the model (SpringSimPy 1.0), additional data were used for model testing: number of eggs produced per individual over its entire life span, from several studies. Only one of the validation studies includes toxicant stress, albeit perhaps a less relevant one from the PPP perspective (copper, constant exposure), and only in a rather short-term laboratory setting (a standard 28-day test). The ability of the model to capture toxic effects on the individual life cycle or on the population dynamics (in heterogeneous environments) is thus not clearly demonstrated in these studies.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

Data quality is not discussed, references to the sources have been provided.

Is the dataset relevant in view of the problem definition?

Yes, but the data sets used are quite limited. See above.

Is the fit of model output to the data good enough?

The correspondence of the model to these data sets was fair, although the data basis was limited. References are provided for the data used, but their quality is not discussed. Furthermore, it is not entirely clear whether and how the model scenario was matched to the conditions of the validation data sets, e.g., in terms of the food scenario. The model output in these comparisons is different in the last TRACE document from those in the earlier ones, which is not explained. A number of changes have been made to the model since the first version, which are likely responsible for these differences.

Has the performance of the model been reported in an objective and reproducible way?

Yes.

2.7.3.10 Model Use

Is a user manual available?

No user manual is available at this moment (in preparation by RIFCON), but extensive documentation (ODD and TRACE) is publicly available for the first version of the model, providing a lot of information (in a structured format) on the model concepts and its parameterisation and testing. The model is now owned by RIFCON, and the software and documentation for the last version are attainable by authorities upon request. RIFCON has produced a version in Python with an updated and extended TRACE document, which is more complete and better structured than the previous versions.

Have all aspects of the modelling cycle been documented?

The documentation mentioned in the previous point seems to cover most elements of the 'modelling cycle', at least those that led to the analysis in the publication to which they belong.

Has a summary sheet been provided by the modeller?

This item cannot be answered. A summary sheet would need to be supplied when the model is used for a specific dossier.

2.7.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

The model is currently owned, maintained and further developed by RIFCON (contact info@rifcon.de), who also offer model analyses with this model to clients.

Is a version control system implemented?

The model has undergone a number of changes since the first version by Meli et al. (2013); these changes have not been subject to version control. With the move to the Python implementation, a formal version control (SVN) is now in place.

2.7.3.12 Overall Judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

It is difficult to answer this question as it is unknown how the model will be used in the regulatory setting, and what the demands are for models to be useful in this respect. No model can be expected to provide accurate and precise predictions of population dynamics under various forms of stress in the field (and the same holds for the methods that are currently used in ERA). However, models may shed more light on several aspects of the lab-to-field extrapolation, and allow running what-if scenarios to identify potential problems.

Strength of the model is the inclusion of spatial heterogeneity and the seasonal influences. However, it only considers a single species, and the representation of the individual's life history (and the effects of stressors on life history) in the model is rather simplistic, and includes a large range of descriptive rules. Even though these rules are generally pragmatic and reasonable, it is difficult to judge whether they together provide an adequate representation of the individual's life history at different food levels, temperatures and toxicant levels (and the interactions between these factors).

Several of the rules are more *ad hoc*, and more questionable. For example, the 'energy budget' is not influenced by age (or stage), temperature or toxicant stress. Further, every individual obtains a new set of randomly-drawn parameters at the start of each season or each day, and each individual will thus change instantly at this point (independent from its state on the previous day). It is not clear to what extent these pragmatic rules influence the realism of the population projections. The model does produce reasonable patterns at the population level, which are quite consistent with the (very limited) population observations, as demonstrated in a few validation cases. As several individual-level parameters were calibrated to population patterns, there is a risk is that the model provides a good description of population statistics for the wrong reasons. The current level of validation thus provides little support for the usefulness of the model.

The toxic effects are represented by static dose-response relationships. This is terribly crude, but it should be realised that the standard toxicity test with *F. candida* is rather useless for parameterising more realistic models for the toxicant effect, as observations are made at the end of the test only. So far, it is unclear whether the model can provide an adequate description for the life cycle of an individual as a function of food availability, temperature and toxic stress (and movement), especially as a number of logical interactions were not included, such as the effect of temperature and life stage on the energy budget (feeding, maintenance costs, etc.) and the effect of food and toxicants on maturation time.

The model has only a very small user community (and this is unlikely to change as the model is not made publicly available). The four published papers are essentially the result of the work of a single scientist (Mattia Meli, working at that time at Roskilde University). In the published case studies, the model has only been applied to copper sulphate and a hypothetical PPP. The model is currently maintained by RIFCON consultancy, who is further developing the model and also applying it for regulatory purposes.

The model is potentially suitable for regulatory purposes, to explore effects of toxicants and spatial heterogeneity on populations of *F. candida*, and to assess the potential for population recovery after toxic insult. However, due to the lack of interactions with other species, and the crudeness of the individual's representation and the food dynamics, the results should be carefully weighed in regulatory decision making. For comparison, in aquatic ERA the recovery of populations can be considered only if assessed in a community context. More validation work with more detailed data sets at both the individual and the population level would be needed to increase confidence in the model. In the current version of the model, there are several worst-case assumptions worked in, and worst-case scenarios can be run to make the overall model analysis conservative.

2.7.4 Qualitative Assessment of Uncertainties

The structure of the SpringSim model is not by itself particularly conservative or non-conservative. The level of conservatism mainly depends on how the model is used, which data are used to calibrate it, which model outputs are used, what scenario is run, etc. We can, however, identify certain areas that will affect the level of conservatism, which are presented below. This level of conservatism has to be considered in an overall risk assessment.

2.7.4.1 Potential for Underestimation of Real Risk

- The basic model considers a population in a closed environment with a regular supply of food. As a consequence, the population will grow to a carrying capacity which is determined by the food level and its re-stocking scenario (and the crowding effects). In such a setting, various types of toxic effects on the individual will hardly affect equilibrium population density. For example, even strong toxic effects on individual reproduction may disappear in the equilibrium situation as there is very little reproduction anyway, close to the carrying capacity. However, under different conditions, the effects may become much more important, e.g., when the population is kept in check by predators or when the population is responding to a seasonal increase in food or recovering from a toxicant pulse. There is thus potential for underestimation of risks (depending on how it is defined), but this is mainly a matter of carefully selecting appropriate a range of realistic scenarios and the appropriate model outputs.
- No effects of toxicants on maturation time, number of broods, 'energy budget' or movement are considered. However, several of these factors will also have occurred in the toxicity test used to parameterise the dose-response relationship. Therefore, their effects are partly included into the model simulation through the reduction in brood size. It is good to note that current standard tests do not allow for parameterising a more realistic toxicity module.
- The model considers a population of *F. candida*, living in isolation. No additional stresses such as predation, disease, competition with other species, or indirect effects through the food source (e.g., when considering fungicides) included. It is likely that these additional stresses will mainly (though not necessarily exclusively) increase the impacts of chemical stress. Furthermore, in terms of its life history (parthenogenetic, opportunistic, capable of rapid population growth), *F. candida* does not seem to be very vulnerable in an ecological sense. It is possible that populations of other collembolan species will respond more strongly to the same exposure situation.

2.7.4.2 Potential for Overestimation of Real Risk

- Food availability is (by default) set rather low, at 5 or 10% of the grid cells. Low food availability likely makes the population more vulnerable to toxic effects. The food dynamics are, however, rather *ad hoc* (food is restocked the next day after it is empty).
- The animals are assumed to be restricted to the top 5 cm of the soil, experiencing maximum exposure to the chemicals (no vertical movement considered). However, the degree of overestimation of risk depends on how the (homogeneous) concentration in that top layer is calculated.
- No possibility for recolonization due to migration from unpolluted areas outside the modelled section.
- In the latest version of the model, eggs are not explicitly included, but rather are 'attached' to the mother. When the mother dies between the moment of egg deposition and egg hatching, the eggs die as well.

2.7.4.3 Potential for Uncertainty in Either Direction

- There is no explicit consideration of bioavailability. Bioavailability of chemicals will generally be different in a field situation than in a toxicity test. For organic chemicals, bioavailability can to

some extent be included by compensating for the difference in organic-matter content between the field soil and the soil used for the toxicity test. However, ageing (sequestration) may reduce bioavailability of chemicals in the field over longer time scales. Bioavailability has to be dealt with separately from the effects modelling.

- The toxicity module in the original version (as used for the published papers) applies a time correction for the effect according to the number of hours spent on contaminated cells. This will underestimate effects for fast-acting compounds, but also for slow-acting ones when the time spent on contaminated cells exceeds 28 days (effects are maximised to the effects observed in the 28-day test). This module has been changed in the current version to a generally more worst-case one: immediate effects according to the 28-day dose-response curve. A static dose response will overestimate the onset of effects, but also overestimates recovery when the individual moves to a clean patch or when the chemical disappears from the soil. Therefore, the level of conservatism will depend on the exposure scenario (i.e., how the chemical is distributed in the environment), and on the TKTD properties of the chemical in this species. The toxicity module can thus lead to errors in both directions. It is good to stress that current standard tests do not allow for parameterisation of more realistic toxicity modules.
- The model is likely very sensitive to the incorporation and strength of the avoidance effect. If (too much) avoidance is assumed when the chemical is not sensed by the organisms, this leads to underestimation of risk. If no (or insufficient) avoidance is assumed while the chemical is sensed, overestimation will result. In the last version, avoidance of chemicals was turned off, which can be considered a worst-case assumption.
- No interactions between food limitation, temperature, crowding and toxicant stress (all factors are assumed to act independently).
- The individual's 'energy budget' is crude, with unclear impacts on the model output. Reproduction is affected by the 'energy status' of the organism and by crowding, but reproduction does not cost any energy, and crowding does not affect the energy budget. No effect of age (or stage) on the energy-budget processes. Juveniles feed as much as adults, and also pay the same costs for maintenance and movement. Parameters of the energy budget are not based on information about the organism's life history or behaviour; they are calibrated to produce a reasonable match to two observed patterns at the population level. It has not been demonstrated that the individual module can provide an adequate representation of the life cycle (and the interactions with the various stressors).
- For each life-history trait, individuals draw a new value from a distribution at the start of a new season (original version) or a new day (last version). Even though the ranges are based on empirical data, it is unclear whether they should be taken independent (traits may well be correlated). Further, it is likely that an individual with a low value for trait A will keep a low value also in subsequent time periods.
- No effects of food limitation or crowding on the maturation time or on the total number of broods.

2.8 eVole

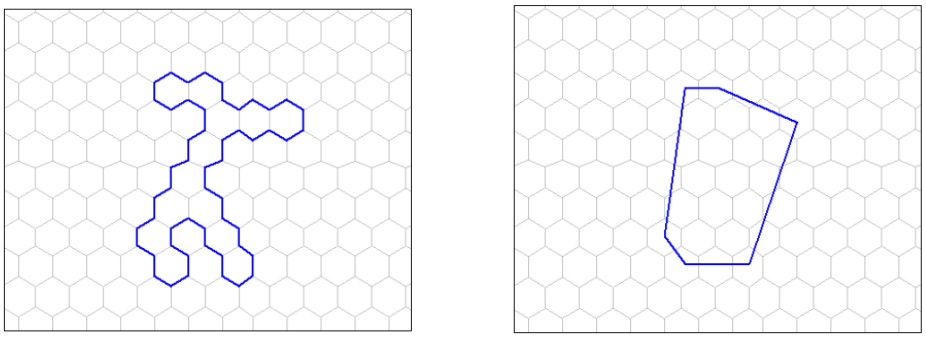
Evaluation by Jeremias Becker and Mathias Franz

2.8.1 General Information

2.8.1.1 Background and Concept

The RIFCON Common Vole Population Model, also called eVole, is a spatially explicit individual based model that is designed for small territorial mammals. The model was first developed and published by Wang and Grimm (2007)[14] for the common shrew *Sorex araneus*. Wang (2013)[15] adapted the model to the common vole *Microtus arvalis* and set it up for the risk assessment of plant protection products. A TRACE documentation of the more recent version 3.0 that contains also a built-in toxicity module for individual-level effects has not been published but was available for this evaluation (RIFCON 2018)[16]. A specific feature of this model is the explicit description of home range dynamics of individuals, from which the population dynamics emerge indirectly. The main motivation for the development of this model was the argument that population dynamics of territorial animals are likely to depend strongly on the availability of suitable habitats for individuals, i.e. the size, shape and location of home ranges in a landscape (Wang and Grimm 2007). While most existing models assume static home ranges, this model explicitly considers how home ranges may change over time. The implicit assumption seems to be that the inclusion of home range dynamics allows for better predictions of population dynamics and thus makes the model more suitable for risk assessment applications.

Figure 11: eVole – Representation of Home Ranges



Left: The home range of an individual consists of hexagonal landscape cells with 5 m diameter. Right: For visualization and comparison with telemetric field data, home ranges were simplified in output graphs by connecting the outermost cells to minimum convex polygons. Graph obtained from RIFCON (2018).

The model uses a landscape, modelled as a grid of hexagonal cells with 5 m diameter each, which consists of different habitat types. Each habitat type provides a (seasonally changing) food value (model for *S. araneus*) or vegetation height and vegetation cover (model for *M. arvalis*) per landscape cell that are the main drivers of home range dynamics. Vegetation height is considered as a proxy for food availability, vegetation cover as a proxy for protection from predation. Individuals dynamically add or remove areas in order to optimize the total food availability in their home range, given that sufficient

[14] Wang, M. and V. Grimm (2007): Home range dynamics and population regulation: An individual-based model of the common shrew Sorex ayaneus. Ecological Modelling 205(3-4): 397-409.

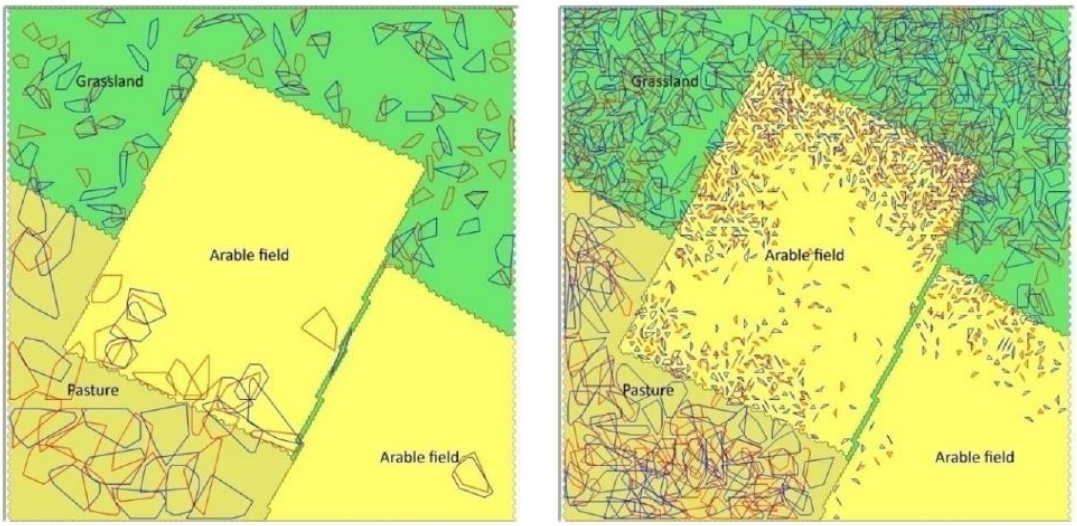
[15] Wang, M. (2013): From home range dynamics to population cycles: Validation and realism of a common vole population model for pesticide risk assessment. Integrated Environmental Assessment and Management 9(2): 294-307.

[16] RIFCON (2018): TRACE document for eVole - RIFCON Common Vole Population Model (Version 3.0). RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany.

cover is available. In this process social criteria such as the presence of other voles are of influence where e.g. adult females may aviod home range overlap (territorial) while adult males during the breeding season seek home range overlap with a potential mate (weaned juvenile female or adult female). Individuals with an insufficient amount of food and / or cover in their home ranges disperse in search of a new home range. Thus, the size, shape and location of home ranges emerge from underlying behavioural rules. Effects of predators, interspecific competitors or of abiotic factors other than food and cover on home range dynamics are not explicitly simulated. Population dynamics in the model emerge from rates for survival and reproduction that are specific for age-classes and sexes and may be influenced by dispersal. Density regulation is not imposed via density-dependent survival rates but instead emerges from home range dynamics: higher population sizes lead to an increasing number of dispersers which do not reproduce as long as they do not manage to establish a home range; additionally, gestating females abort their litter upon dispersal, and the offspring of a lactating female die when that female disperses. Dispersers can be, as an option in the model, configured to have also lower survival rates.

In its original version (Wang and Grimm 2007), the model was not yet specifically set up for the risk assessment of pesticides. Thus, applications in the context of risk assessment needed to add external sub-models that try to capture the effects of toxicants on the individual level, e. g. on survival rates and on reproduction. Recent versions (RIFCON 2018) contain a built-in toxicity module that estimates the dietary dose for each individual based on the location of its home range and exposure equations proposed for Tier 1 in the respective EFSA guidance document (EFSA 2009b). Based on the exposure each individual can be subjected to lethal and a number of sublethal effects using customized dose-response relationships (see model description below). However, eVole does not simulate the body weight of individuals and how effects on body weight might impact the performance (e. g. spatial behaviour and competitiveness). Decreasing body weight is a sensitive endpoint in many toxicological studies on small mammals. Therefore, in eVole effects on body weight may be converted to increased mortality using a relation reported in a semi-field study from Oksanen et al. (2007).

Figure 12: eVole – Seasonal Variation in Size and Distribution of Home Ranges



Size and spatial distribution of home ranges at the end of March when food availability is low (left), and in the beginning of August (right) when the amount of food in arable fields is high. Graph obtained from RIFCON (2018).

2.8.1.2 **Status of the Model**

Model for the Common Shrew

The population model was initially developed and parameterized for the common shrew *Sorex araneus* (Wang and Grimm 2007). The predicted population and home range dynamics in the absence of pesticides were successfully validated with field data: The percentage of pregnant or breeding females during the breeding season that emerged from the model (average 83 %, up to 100 % during peak of breeding season) was close to field observations (up to 90 % and more). After calibration of daily mortality rates to sustain long-term population persistence, 60 % of pups died within 2 months in the model, which was comparable to field observations (30 – 60 %). However, the modelled fluctuation in population size between years was lower than in field studies, specifically if no yearly fluctuation in food availability was imposed (in field studies food availability was reported to vary by up to 70 % between years). Seasonal changes in the age structure and spatial distribution of home ranges matched field observations.

Wang and Grimm (2010) applied this model to investigate the usefulness of different modelling endpoints for risk assessment (density, population growth rate, and population viability = risk of extinction within 50 years). For this purpose, the mortality of all individuals was increased by 10 % or 20 % at a single day per year to simulate the effect of a single pesticide application. Population density responded clearly stronger than growth rate, which only shortly decreased in the month of exposure and even increased three months later due to density-dependent recovery. Effects on the population density were not larger, but longer detectable when hedges (providing high food, but also contamination in this case study) were added to the landscape of cereal fields. This observation was explained by the fact that hedges increased the overall population density which in turn decreased stochastic fluctuation in the population size. Thus, the statistical power to detect pesticide effects increased, even when the magnitude of effects remained constant or decreased. Also population viability was a sensitive endpoint, particularly when hedges were removed from the landscape. In the presence of hedges, 20 % mortality at July 15 (middle of breeding season, highest population density) caused a long-term decline in population density over many years leading eventually to extinction. In contrast, the same mortality experienced at April 1 (beginning of breeding season) allowed the seasonal cycles in population density to stabilize at a lowered level after 5 years of application because the young generation was not affected in spring. When hedges were removed, however, also 20 % mortality in spring resulted in a continuous population decline.

Model for the Common Vole

Wang (2013) adapted and re-parameterized the model to the common vole *Microtus arvalis*, with small changes in the code (introduction of vegetation cover, which must be sufficiently high in a cell to be part of a home range, permission of home range overlapping except for adult females). Population and home range dynamics without pesticide exposure in this new model version that was called eVole were also successfully validated with field data (see section 2.8.2.9 of the model description below). However, predictions of eVole on how pesticide effects on individuals propagate to the population level have never been validated. The model is well-described based on the ODD protocol (Grimm et al. 2006). The source code is not freely available but can be provided to authorities upon request, in addition to the model executable and a user manual. If the request is related to a certain model application under evaluation, the simulation files and R scripts used to analyse the model output can be also provided promptly upon request from the authority (RIFCON, personal communication).

Two applications of eVole have been published to demonstrate its potential for use in the risk assessment of pesticides. The studies used early model versions which did not include a toxicity module; therefore individual-level effects of pesticides have been estimated separately and then imposed as

probability distributions to the model. Nevertheless, these applications illustrate the challenges that arise when linking the population model to toxicological data on sublethal effects on individuals.

In the first example, Bastiansen et al. (2013) applied eVole to the insecticide sulfoxaflor which affects survival of newborn pups when exposed during a critical time window during pregnancy. Insecticide applications in May and August were considered. Early versions of eVole did not include a toxicity module for individual-level effects of pesticides, and the toxicity module introduced with version 3.0 may only be used to calculate exposure of individuals that have been born. The exposure of unborn offspring in the womb and the resulting individual-level effect was therefore calculated independently from the model: A dose-response curve for post-natal death of pups and limb abnormalities (also considered as death) vs. maternal exposure during the critical time of pregnancy (measured as DDD) was created from toxicological data. For the maternal exposure during the critical time window, a distribution of daily dietary doses was calculated from toxicokinetic equations using probability distributions for maternal body weight, food energy, moisture, assimilation efficiency, and for the day of the gestation cycle. A mixed diet of grasses and non-grass herbs was assumed. The dose-response function was then applied to the maternal exposure distribution to obtain an effect distribution for the increase in pup mortality (probability of effect vs. magnitude of effect). In eVole, exposed pregnant females then drew a random value from this effect distribution at the time of pesticide application. The effect distribution ranged from ≤ 5 % effect in 25 % of individuals to > 90 % effect in < 0.5 % of individuals. The model predicted only negligible effects on the population density investigated at every 1st January (outside of breeding season).

The second application example comes from the SETAC technical workshop MODELINK: Schmitt et al. (2015) applied three individual-based population models for small mammals (eVole for the common vole, ALMaSS for the field vole and the model of Liu et al. (2013) for the wood mouse) to a hypothetical fungicide in order to assess the suitability of these models for refined risk assessment. The authors considered a scenario in which the hypothetical fungicide failed Tier 1 because the toxicity effect ratio TER < 5, which requires a refined risk assessment to demonstrate that small effects (LOAEL) on the individual level do not cause unacceptable (> 5% or long-term) effects on the population. A single application per year was assumed, but the timing and the magnitude of imposed individual-level effects remain unclear from the publication. The fungicide was considered to affect individual survival, the litter size and the time to first reproduction, using an average toxicological profile across 15 real pesticides. The applied version of eVole did not include a toxicity module yet, therefore again effects at the individual level were imposed as a probability distribution that considers the assumed natural variation in exposure and toxicity. From this probability distribution a random value was distributed to each individual for the pesticide-induced percent decrease in survival, litter size and increase in time to reproduction. The results of all three population models suggested a low risk because the populations quickly recovered from the effects of a pulse exposure due to high reproduction rates (Schmitt et al. 2015). eVole predicted a considerable (> 5%) reduction in the population density at 1st January (outside the application season and before the breeding season) only if the expected application rate was 10x increased or if pesticide-induced mortality was set to 20 % as positive control. Even then, no long-term decline in population densities across 11 years with fungicide exposure was predicted.

The following model description and evaluation focuses on eVole version 3.0 (RIFCON 2018). Additionally, differences to the model version for the common shrew (Wang and Grimm 2007) are described where appropriate.

2.8.2 Model Description

2.8.2.1 Problem Definition

Context in which the model will be used

The RIFCON Common Vole Population Model – eVole – is based on a population model for the common shrew (*Sorex araneus*) published by Wang and Grimm (2007) to capture the relation between home range and population dynamics and to predict effects of pesticides and agricultural practices to small mammals. Wang (2013) re-parameterized the model for the common vole (*Microtus arvalis*) and developed it specifically further for Higher Tier assessments on small herbivorous mammals using the common vole as focal species.

Specification of the question(s) that should be answered with the model

The specific purpose of the model is the realistic simulation population dynamics of the common vole in order to assess recovery from potential impacts on the populations by plant protection products (RIFCON 2018). Considering the risk assessment frame work (Regulation (EC) No 1107/2009), the overall aim is to assess whether effects on individual survival, reproduction or behaviour may result in unacceptable effects on the population size, structure or viability in the field.

Specification of necessary model outputs and protection goals

For mammals, the actual protection goal of "clearly establishing that there will be no visible mortality and no long-term repercussions for abundance and diversity" and/or the more conservative surrogate protection goal of "making any mortality or reproductive effects unlikely" can be addressed (EFSA 2009b). Negligible to small effects on the abundance and population structure for days to weeks have been considered acceptable, but only if caused by avoidance behaviour (EFSA PPR 2010).

In this context, the model may be used to assess whether reproductive effects from poisoning will result in any significant effect on the abundance and population structure, and whether acute or reproductive effects that result from avoidance behaviour will result in long-term repercussions. Addressing these questions will require modelling of the magnitude of effects on the population size and structure in a field to landscape context and the time for population recovery (in days to weeks).

Domain of applicability of the model

eVole simulates the population dynamics of common voles in a seasonal landscape in daily time steps with a spatial resolution of 5 m. The model might be re-parameterized and applied also to other territorial small mammals without fundamental changes in the conceptual model (but would need to be validated again). Climatic or biotic factors other than intraspecific competition are not explicitly simulated. However, seasonality is implicitly accounted for by the vegetation dynamics in the landscape as well as several parameters of the ecological model, specifically natural mortality of weaned juveniles and adults, litter size as well as food consumption. As noted in RIFCON (2018), the domain of applicability is limited to the climatic and community context used for parameterization (Central Europe). Indirect effects and effects from bioaccumulation are not covered, which may be particularly relevant when the model is applied to non-herbivorous species.

Why is the model being used?

When the required TER values in Tier 1 assessment have not been reached, compliance with the protection goals must be demonstrated at population and community level. Since classical field-studies for

Higher Tier assessment are hardly feasible for mammals, it is often necessary to project from individual-level endpoints from toxicological studies to population level effects. Therefore, Schmitt et al. (2015) consider eVole and other population models as an important tool for risk assessment to conduct this projection consistently and transparently, allowing for the emergence of case-specific relationships between the fate of individuals and the populations under consideration.

What protection goal is being addressed?

The specific protection goal depends on a specific model application and cannot be addressed for the model in general. As outlined above, typical protection goals to be addressed with eVole include no (long-term) effects on population size and structure.

What outputs are required?

Addressing the potential protection goals outlined above will require the simulation of population development over time in a seasonally changing landscape, because individuals may move between habitats that differ in terms of quality and exposure in time and space. The most relevant endpoint that can be addressed with the model is the time for population recovery after exposure and the risk of increasing population-level effects across repeated applications.

How was the species chosen?

The common vole (*Microtus arvalis*) has been identified as representative for the generic focal species of small herbivorous mammals ("voles") in the EFSA guidance on risk assessment for birds and mammals (EFSA 2009b). The species is common and feeds primarily on plant material in agricultural fields (Niethammer and Krapp 1982), and therefore experiences high risk of exposure to plant protection products. It is one of the best studied species among small mammals with relatively comprehensive data available for modelling (EFSA 2009b, RIFCON 2018).

Which other species/groups are being covered by the chosen one(s)?

The ecological vulnerability of the common vole is generally low due to its short generation time (33 d in summer) and high reproductive output (Niethammer and Krapp 1982). The common vole is even considered a pest species characterized by inter-annual population cycles with mass development (Truszkowski 1982). However, these traits are shared with the other species covered by the generic focal species of small herbivorous mammals (EFSA 2009b). Therefore, modelling effects on the common vole as focal species in Higher Tier assessments is likely to cover also the other members of the small herbivorous mammals to a large extent.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Model predictions on the spatial arrangement of home ranges and on population dynamics without pesticide exposure have been tested with various field observations from the literature (see Wang 2013, RIFCON 2018). Pattern matching between observed and predicted data has not been quantified and no required degree of matching has been pre-defined, probably because this has not been specified in any guidelines and there is no common convention on the degree of pattern matching. Predicted effects of pesticide exposure have not been validated with real-world observations.

2.8.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

Species-specific parameters were parameterized using data from the literature, in most cases field studies on *Microtus arvalis* in different crops in Central Europe (Wang 2013, RIFCON 2018). This included habitat- and age class-specific survival rates and reproductive behaviour, seasonal changes in home range size, the maximum dispersal distance, and the minimum food demand for maintaining a home range. Both field and laboratory data were consulted for the parameterisation of the duration of the gestation and lactation period, maximum age, litter size, and sex ratio at birth. Development time to maturity of male voles and the minimum vegetation cover required to be part of a home range were taken from the related species *M. agrestis*. However, the seasonally changing food consumption (link between vegetation height and home range size) and the maximum rate of increase in home range size had to be calibrated.

Assessment of the quality of the data

All data used descend from peer-reviewed scientific literature. The model documentation provides a sound review on available data from field and laboratory studies including an assessment of the data quality by the authors (RIFCON 2018). Most parameters that are expected to depend significantly on environmental conditions (such as timing of breeding season) were derived from field studies in various agricultural habitats. Background mortality was derived from mark-recapture studies in which individuals that were not trapped again were considered dead (worst-case assumption).

2.8.2.3 Conceptual Model

Description of the model concepts including a diagram

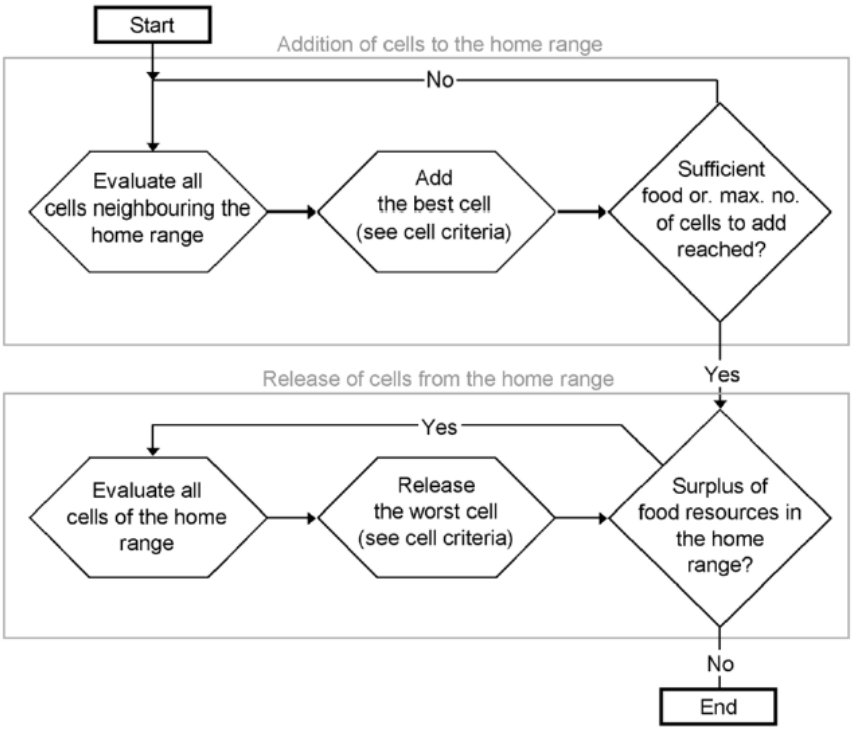
In eVole, population dynamics emerge from reproduction success that is coupled to the seasonally varying amount of available home ranges (Wang and Grimm 2007, Wang 2013, RIFCON 2018). Individuals develop from pup to adult with an imposed developmental rate and background mortality; also the litter size is imposed, i. e. not dynamically affected by population density. However, in order to reproduce, individuals need to establish a home range of connected cells in a landscape grid that meets the individual's food demand. Each cell has a diameter of 5 m and provides a food amount that depends on the habitat type, season and optionally on simulated farming practice (decrease due to mowing events). Population dynamics are therefore indirectly regulated by the landscape composition of habitat types and season (determine food and vegetation cover), and by the population density, all of which determine the number of reproducing individuals. Vegetation cover is a proxy for shelter; only cells exceeding a minimum vegetation cover can be used as part of a home range. The daily background mortality varies with sex, season and age class, and the food demand and litter size vary with the season based on data from field studies.

The size and location of home ranges is driven by daily home range optimization: Individuals prefer cells that provide a high food amount (represented by vegetation height as a proxy and not depleted by animals) and are not occupied by competing individuals. Individuals can sense the amount of food, vegetation cover (shelter) and the presence of other individuals (competition and search for partner during the breeding season) in each cell of their home range and in all neighbouring cells. Based on this information, each individual identifies the most suitable cells; preferences vary with the sex, the developmental stage and the season.

Adult females may dispel each other from occupied cells due to territoriality (winner is chosen randomly). The maximum number of cells that can be added to a home range each day is limited. If the

overall food amount in a home range surpasses the food demand of an individual, least preferred cells at the border of the home range are released until the food demand is met. If the food amount of a home range falls below a threshold and the home range is not currently increasing, individuals disperse in search of a new home range. Dispersers do not mate and an increased mortality for dispersing individuals can optionally be incorporated. When the home ranges of an adult male and female overlap during the breeding season, mating is considered successful and the female becomes immediately pregnant, followed by a gestation period before the female can mate again (in the version for the common shrew females wait until the end of lactation before mating again). In the breeding season, adult males therefore prefer cells that overlap with those of a female's home range and disperse when their home range does not overlap with any home range of a female.

Figure 13: eVole – Flowchart of Home Range Optimization



Algorithm for home range optimization in eVole. Individuals evaluate and add most preferable neighbouring landscape cells until food demand or the maximum home range increase per day is met. In case of a food surplus, individuals afterwards release least preferable cells until food demand is met. Graph obtained from RIFCON (2018).

Identify the main components and processes in the system

The main components are a landscape and the individuals with their associated home range. The basic processes comprise home range dynamics (incl. foraging, intraspecific competition and search for mating partners), dispersal, reproduction and survival. These basic processes are influenced by the landscape (food amount and vegetation cover determine the number of available home ranges that can be established and thus the number of reproducing individuals).

How the effects of the chemicals are modelled

eVole may be coupled to an external module for individual-level effects. Additionally, since version 3.0 also a built-in toxicity module for individual-level effects has been available (RIFCON 2018). For this module, the user needs to enter the application rate and pattern of a substance, its residue unit dose (RUD) in vole food, its DT50 value, and the dose-response curve parameters for the relevant effects.

eVole then calculates the foliar residues in vole food for each landscape cell on a daily basis, considering substance dissipation based on first-order kinetics. The dietary exposure of each vole depends on the foliar residues and the intake: The dietary uptake from the different cells of a vole's home range is thus proportional to the food resources in each cell. The dietary exposure (mg substance/kg body weight) is calculated according to the guidance for risk assessment on birds and mammals (EFSA 2009b) based on foliar residues, body weight, food energy content, food moisture content, and the assimilation efficiency of food. Default values are provided for all the parameters except foliar residues. Each day, dietary exposure of each individual is quantified as the current daily dietary dose (DDDt), as the long-term dietary dose (time-weighted average DDDt across the last 21 d, TWA21), and as highest DDDt during the last 21 days (MAX21). Additional exposure routes such as dermal and respiratory exposure during spraying are not supported.

The dietary exposure is related to an individual-level effect by specifying either a log-logistic or probit dose-response function or by providing data points for linear interpolation. The functions provide a dose-dependent threshold; every day, each individual draws a random number between 0 and 1 and the effect is executed if the number is lower than the current threshold. With data from acute (DDDt) or chronic (TWA21) tests, each of the following effects can be linked to dietary exposure (DDDt, TWA21 or max. MAX21), using separate dose-response curves:

- Direct mortality (additional daily probability of death, can be separated for sexes and developmental stages)
- Reduced litter size (each offspring individual of a mother has a certain probability to die at the time of birth equal to the percentage of litter-size reduction)
- Indirect mortality as consequence of reduced body weight (because the main population model does not consider body weight, effects on body weight may be translated to increased daily background mortality using an implemented relation found in a semi-field study on the bank vole *Myodes glareolus* (Oksanen et al. 2007)
- Abortion of litter (daily probability for pregnant female to lose all offspring)
- Increased gestation period (here the dose-response function ranges from 0 to the max. possible delay [d], and each individual draws a random number from that range; the gestation period increases by the integer part of that number in days, and with the probability provided by the decimal part another day is added)
- Increased F1 mortality (additional daily probability of death for suckling juveniles, based on their mother's exposure)
- Delayed F1 maturation (increase in the maturation time established at birth, based on maternal exposure; calculated in the same way as increased gestation period)

How the components and processes are linked

Behavioural rules and the population density determine individual home range dynamics and dispersal. Home range dynamics and dispersal influence reproduction (because dispersers cannot mate, a gestating female that disperses aborts its litter and the sucklings die when their lactating mother disperses) and mortality (for the common shrew and optionally also for the common vole). Mortality and reproduction influence the population density, which again determines home range availability. In contrast to reproduction, daily mortality is static (not affected by emerging properties such as population density); therefore, population dynamics in the absence of pesticides are mainly driven by reproduction in this model. Pesticides can affect mortality, reproduction and developmental parameters in the model, which propagates to effects on the population and home range dynamics.

2.8.2.4 **Formal Model**

Identification of the model variables

Individuals are characterized by the following state variables: age, sex, developmental stage (suckling offspring, subadult and adult), sexual activity (breeding / non-breeding, defined by breeding season), fertility (individuals are infertile when they are immature and during gestation), maternal state (pregnant, lactating), the occupied home range and the actual daily and long-term dietary dose. Home ranges are represented by a number of connected hexagonal landscape cells of 5 m diameter used by an individual. Each cell is characterized by its position, habitat type (determines current vegetation height and cover), vegetation height (seasonally changing food value), vegetation cover (seasonally changing shelter, a minimum is required to be part of a home range), the presence of home ranges, and by pesticide exposure.

Identification of the model parameters

15 parameters related to life history traits (Wang & Grimm 2007, Wang 2013) describe the mortality (daily background mortality, daily dispersal mortality (only for common shrew), maximum age), reproduction (start/end of the breeding season, timeframe when an individual is fertile, time lag before females become fertile after beginning of the breeding season or after giving birth, gestation length, lactation length, litter size, sex ratio at birth, time to sexual maturity), home ranges (max. number of cells added to a home range each day, food demand (varies seasonally and between sexes for individuals that possess a home range) and dispersal (maximum dispersal distance per day, dispersal threshold = minimum food availability (in % of food demand) before dispersal is initiated). Values for the majority of parameters are drawn from probability distributions in order to reflect natural and / or individual variability.

Description of the most important model equations or algorithms

The simulation proceeds in time steps of 1 day. The key algorithm is the optimization of home ranges. Individuals replace cells with low food by cells with high food and generally prefer cells not occupied by another individual (except for males during the breeding season which prefer cells that are part of a female's home range). If food available in a home range is below the demand of an individual, the individual will try to increase its home range. If the food resources surpass the demand, then the individual will release cells from its home range. The importance of intraspecific competition in the optimization procedure depends on the developmental stage. In the version for the common shrew, individuals of the same developmental stage have a similar chance to take over a cell. Adult females always expel subadults, forcing subadults to disperse from their mother's home range. Adult males do not compete for cells during the breeding season, since their home ranges are reported to overlap largely with the home ranges of other individuals of both sexes (Wang & Grimm 2007). In the version for the common vole, only adult females expel each other. Otherwise home ranges can overlap, although individuals prefer empty cells.

To avoid gaps within home ranges and splitting of home ranges, some additional rules have been introduced: (1) only cells at the border of a home range can be released; (2) individuals can be expelled only from cells at the border of their home range; (3) when a home range is split in two or more parts, the expelled animal retreats to the "better" part of the home range; (4) cells enclosed by the addition of a new home range cell are included in the home range. Individuals become dispersers when they are expelled from their home range or when the food availability of the home range falls below their dispersal threshold (critical percentage of food demand). Adult males become also dispersers during breeding season when their home range does not overlap with a female. Dispersers move in a random

direction each day until a maximum distance is reached or until an appropriate cell is found to establish a new home-range. If an appropriate cell is found on the way (food, shelter, no competing individuals, the presence of mating partners in case of fertile males) it is used as the first cell of a new home range. During the breeding season, a fertile female and male mate when their home ranges overlap. Afterwards, time of birth is determined by the length of the gestation period. The update of the developmental stage of the offspring is based on the length of the lactation period and the time until maturity.

Table 9: eVole – Quality Criteria for Use of Landscape Cells as Home Range

Adult males	Adult females	Weaned juveniles (both sexes)
Breeding season		
Presence of adult males (-10^9)	Presence adult females (-10^9)	Presence any individuals (-10^9)
Presence of adult females (10^{12})	Food resources (1)	Food resources (1)
Food resources (1)	Presence juveniles (-10^3)	
Presence of weaned juveniles (-10^3)		
Non-breeding season		
Presence of any individuals (-10^9)	Presence any individuals (-10^9)	Presence any individuals (-10^9)
Food resources (1)	Food resources (1)	Food resources (1)

Criteria for the evaluation of landscape cells for each sex and life stage of the common vole Weights for each criterion are given in parentheses; cells with highest positive weight are preferred. Data from RIFCON (2018, edited).

2.8.2.5 Computer Model

Description of the model implementation

The model has been implemented in C++ (RIFCON 2018) but this is not explicitly mentioned in the TRACE model description. The code is owned by RIFCON GmbH and is not publicly available but can be provided to authorities on request.

Checking the computer model for errors, bugs and inconsistencies in the code

The software implementation was verified by a thorough peer review of the code and comprehensive testing of the model behaviour (Wang 2013, RIFCON 2018).

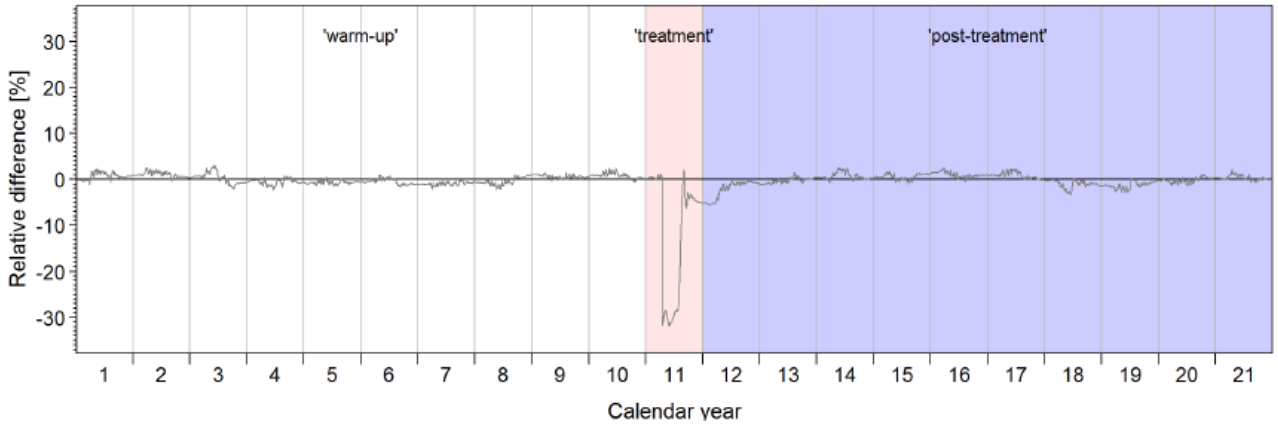
Demonstrate that the computer model performs as indicated by the conceptual and formal models

Verification of the model implementation included visual debugging (e. g. tracing the movement behaviour or fate of individuals), recording values of state variables during test runs, and statistical comparison of model output with expectations from the formal model. Additionally, sub-models for the dissipation of foliar residues and for the dietary uptake were re-implemented in Excel and in R and the output was compared. The procedures have been thoroughly described in the TRACE document (RIFCON 2018).

To demonstrate that negative population effects can be actually produced in eVole, a “positive control” simulation was run in which all individuals experienced 25 % additional mortality on 20th April after

model pre-run for 11 years (RIFCON 2018). The imposed additional 25 % instantaneous mortality resulted in higher short-term decrease of the population (31 %) due to the death of pregnant females which would have given birth to offspring (Fig. 14). Population density recovered within the same breeding season, but dropped again and remained slightly lower than the control density until the beginning of the following breeding season, probably due to demographic effects.

Figure 14: eVole – Demonstration of Generating Effects on Population Density



Comparison of mean population density from 50 control model runs without pesticide effects and from 50 "positive control" simulations with 25 % additional mortality for all individuals at 20th April in year 11. A value of -31 means that the population density in the "treated" scenario is on average 31 % lower than in the control scenario without pesticide effect. Graph obtained from RIFCON (2018).

Additionally, Wang (2013) showed that populations in the model recovered from up to 50 % instantaneous mortality experienced by all individuals at 1st of April within the same breeding season. Populations were considered recovered when no statistical difference in population density was observed between 10 simulations with increased mortality and 10 control runs. In the same study, populations recovered within 1 month from up to 70 % reduction in litter size experienced over 1 week. When reduction in litter size lasted for 2 weeks, more than 10 % reduction resulted in effects that lasted until the following year.

2.8.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context in which the model is run

eVole has been parameterized to simulate a typical population of the common vole in agricultural fields of Central Europe. This refers to those parameters describing the survival, development and behaviour of individuals, including the food demand (that relates home range size to vegetation height) and the need for shelter (that relates home range location to vegetation cover). For the landscape composition, no default settings have been presented.

Publications to demonstrate the model applicability to the scientific community include the studies of Bastiansen et al. (2013) and Schmitt et al. (2015) (see introduction above). In both studies, only treated area was considered to be exposed (no drift). Bastiansen et al. (2013) used a landscape of 75% wheat and 25% surrounding grassland. Schmitt et al. (2015) simulated a landscape of 25 ha with 90 % wheat and 10% surrounding grassland.

Based on the protection goals for mammals, the model is mainly intended to extrapolate from experimentally observed sublethal effects on individuals to potential population-level effects. Common sublethal effects include reduced body weight and developmental abnormalities. However, eVole does not

simulate body weight or related indicators of the physiological state. Therefore, translating experimentally observed effects into effects that can be imposed via the toxicity module is challenging. This case-specific translation may be also considered as part of the environmental scenario. See the introductory text for examples in the case studies of Bastiansen et al. (2013) and Schmitt et al. (2015).

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

In the applications of Bastiansen et al. (2013) and Schmitt et al. (2015), the authors considered the modelled landscape as a simplified worst-case scenario because it represents a consolidated landscape with large continuous cereal fields and only few non-contaminated hedges. However, Schmitt et al. (2015) point out that further investigation and agreement are needed to develop recommendations for landscape attributes such as size, structure, and crop rotation to define appropriate regulatory risk assessment scenarios. The default parameter settings for the individuals were used because they were considered representative for central European populations. No data have been published that would justify parameter adjustment to different climatic conditions, e. g. in the EU Northern Zone.

2.8.2.7 Parameter Estimation

Description of the model parameter estimation

Parameters for the life history and behaviour of individuals are derived from literature. They are either constant values or drawn from random distributions. As an exception, the food demand and the maximum increase in home range size per day have been calibrated due to a lack of data from the literature. Additionally, for the common shrew the background mortality rate and the time lag between gestation and the following mating has been calibrated. See Tab. 1 in Wang (2013) for an overview of parameter values and sources for the common vole, and Tab. 1 in Wang and Grimm (2007) for the common shrew.

Parameters estimated from the literature — what are the sources and why are these appropriate?

Literature used for the parameterization of species traits comprises mostly field studies in different crops in Central Europe (Wang and Grimm 2007, Wang 2013, RIFCON 2018). This includes background (natural) survival rates (constant for sucklings, seasonal and sex-specific for weaned juveniles and adults), seasonal litter size (field specimens) and reproductive behaviour, increased mortality during dispersal (for the common shrew, also optional for the common vole), seasonal and sex-specific (common vole) or age-class specific (common shrew) home range size, the maximum age and dispersal distance, the minimum needed vegetation cover for voles to be present in an area, and the minimum food demand for maintaining a home range. Both field and laboratory data were consulted for the parameterization of the duration of the gestation and lactation period, maximum age, litter size, and sex ratio at birth. A sex ratio at birth of 1:1 was assumed. Development time to maturity of male voles and the minimum vegetation cover required to be part of a home range were taken from studies on the related

Parameters obtained from calibration — how and why this was done?

For both the common shrew and the common vole, the food demand had to be calibrated due to missing data from the literature (Wang and Grimm 2007, RIFCON 2018). This parameter relates the average home range size to the average food amount (represented by vegetation height in the common vole model, and by food biomass in the common shrew model) in a landscape cell. The food demand was calibrated to produce average home range sizes that matched observations from the literature.

This was done separately for males and females. For the common shrew, home range sizes were available for subadults, but sizes were assumed to increase two (females) to four times (males) while breeding, based on information on *Sorex vagrans* (Wang and Grimm 2007). For the common vole, average home range sizes across the year were available from Jacob (2000) and rescaled in midwinter and midsummer by a factor from Reichstein (1960) for the seasonal variation in home range size. Home range sizes in the remaining months were estimated from linear interpolation between winter and summer sizes. Additionally, the number of cells which can be added to a home range per day was calibrated such that the rate of increase for home ranges size was considered realistic by the authors.

For the common shrew model, background mortality rates (not caused by pesticides) were available for the related species *Sorex vagrans* from mark-recapture studies. However, the rates resulted in population extinction after a few years and were considered too high by the authors because all individuals that were not recaptured were considered dead which is probably not realistic. Instead, background mortality was calibrated to produce a mean population growth rate close to zero. The calibrated survival rates were 6 – 25 % higher compared to those from field studies (Wang and Grimm 2007).

Additionally, in the common shrew model simulations with reproduction parameters from the literature resulted in unrealistically high numbers of pregnant females and offspring. Therefore, the authors suggested that females do not mate again directly after giving birth and introduced a time lag before the following reproduction cycle. The length of the time lag was calibrated to yield on average two litters per female per breeding season. For the common vole this time lag was removed.

2.8.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

A local sensitivity analysis has been conducted for the basic model that does not include the sub-model for the effects of chemicals (Wang 2013). For the common vole, this analysis was carried out by varying all input parameters within a range of ± 30 % in steps of 1% and carrying out a 1-year simulation per parameterization, i.e., 61 simulations per parameter. The analysis was carried out in a 25ha mixed landscape with grassland, pasture and arable fields (RIFCON 2018). Regression coefficients were calculated for each parameter using linear regression. This analysis revealed that population density, the production of offspring and population growth were most sensitive to the length of the breeding season, but also to gestation length and adult background mortality (Tab. 26 and 27 in RIFCON 2018). Home range size was mainly affected by the amount of food available in landscape cells. However, variation in the model output was always smaller than the variation in any model parameter, indicating that there was no surprisingly highly influential parameter (RIFCON 2018)

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

Such an analysis has not been conducted and is case-specific, therefore most appropriate once the model has been set up for a specific pesticide and protection goal.

2.8.2.9 Comparison with Measurements

This section deals with the eVole model for the common vole. See section 0 in the general information for tests on the structural integrity of the earlier model for the common shrew.

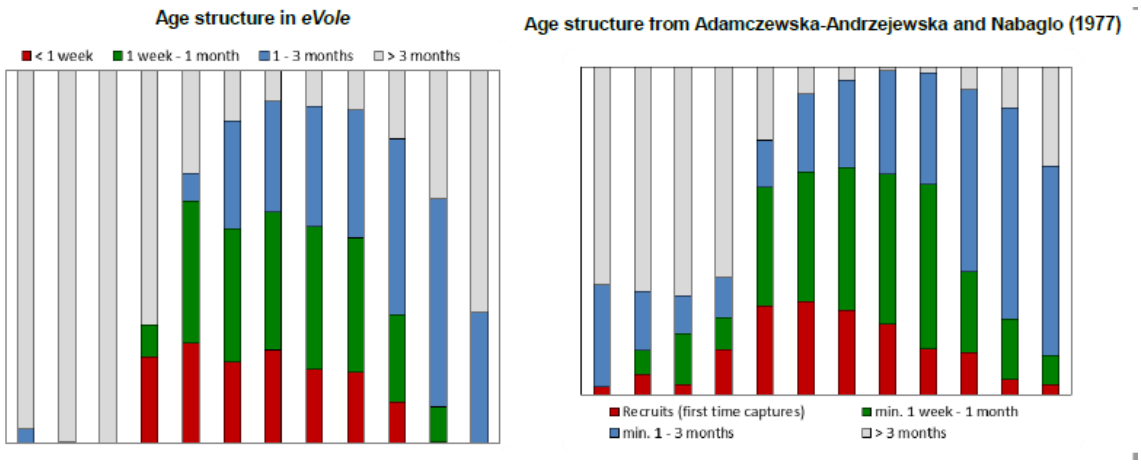
Description of comparisons of model output with independent data

Model predictions on the basic population dynamics of the common vole without pesticide exposure have been validated with real-world observations from field studies (Wang 2013, RIFCON 2018). However, predictions on population effects of pesticides have not been validated. 50-year simulations and 50 1-year simulations were run to compare the simulated reproduction, survival, spatial behaviour and population dynamics to field studies. Some of the studies have been already used for parameterization, but the studied endpoints differed from the input data. To enable comparisons with the field data, the landscape and the initial conditions were matched to the field studies.

Demonstration that the model output provides an adequate match to data patterns

Predictions on reproduction, age structure, spatial behaviour and population dynamics in the absence of pesticides were generally comparable to real-world observations (Tab. 4 in Wang 2013, Tab. 28 in RIFCON 2018): While the litter size is an input parameter, the actual number of offspring per breeding female and lifetime emerged from the model (1 – 6, mean = 1.8) and matched field observations (mean = 2, max = 4, in captivity up to 33). The modelled percentage of pregnant females was relatively close to observations, both in August (33 – 86 (mean = 75) % predicted vs. 60 – 100 % observed) and in September (10 – 71 (mean = 57) % predicted vs. 30 – 50 % observed).

Figure 15: eVole – Validation of Simulated Age Structure



Comparison of the simulated age structure in a pesticide-free scenario (left) and field observations from Adamczewska-Andrzejewska and Nabaglo (1977). Graph obtained from RIFCON (2018).

The age structure emerging in the model generally matched observations during the breeding season, though in January – March, eVole predicted no young individuals (< 3 months) while they were observed in the field. The modellers noted that the mark-recapture field study reported only minimum ages, considering all individuals captured for the first time as recruits. However, the field study showed also a number of re-captured young individuals, suggesting that reproduction took also place outside the breeding season when eVole assumes no breeding. In the simulations, populations reached their seasonally maximum size in July – September which was comparable to field observations (July – October). While the maximum dispersal distance is a pre-defined parameter, the average dispersal distance that emerged from simulations was lower (143.9 m) and matched field observations from the literature (100 – 200 m. Similarly, the model was able to reproduce seasonal changes in the spatial distribution of individuals across habitats (grassland vs. arable fields) due to the seasonally varying food availability (Wang 2013, RIFCON 2018).

eVole was also able to reproduce inter-annual fluctuation in abundance observed in field studies (Wang 2013): When tracking only a small part of the landscape of 0.5 ha that matched the studied area

in the field, population size in a 50 year simulation varied stochastically and could double or half from one year to the next, which matched the variation in field data. The model further produced population cycles of 3 – 4 years length that matched long-term field observations. Tkadlec and Stenseth (2001) observed that population cycles in Central European voles increase toward the South and hypothesized that an extended breeding season in warmer countries stabilizes these cycles. Extending the breeding season in eVole confirmed the emergence of more pronounced population cycles, though the ecological mechanism behind these cycles in real populations (intrinsic or extrinsic from predation) is still under debate.

2.8.2.10 **Model Use**

Explanation of how the model conforms to the requirements set in the problem definition

The authors consider the model fit for regulatory purposes, stating that it has been thoroughly tested and validated based on a pattern-oriented modelling approach. The authors point out that the model has been parameterized and validated with several independent data sets, which avoids the risk of overfitting. Overfitted models might reproduce a specific data set without capturing the appropriate mechanisms and therefore fail when applied to different data within its domain of applicability. The authors suggest to use eVole for the simulation of population dynamics of the common vole in order to predict recovery from potential impacts on the populations by plant protection products (RIFCON 2018).

Description how the model works (user manual)

A detailed model description based on the ODD protocol (Grimm et al. 2006) has been published (Wang and Grimm 2007, Wang 2013) and is part of the TRACE document (RIFCON 2018). A user manual together with the model is available to authorities upon request.

Description of the pesticide parameter values used in the model

The selected mode of actions offered in the toxicity module and the pesticide parameter values are case-specific and therefore not part of the general model. Bastiansen et al. (2013) applied eVole to sulfoxaflor to demonstrate its applicability for risk assessment; in this study toxicological data on postnatal death and limb abnormalities following exposure during gestation were translated to increased pup mortality using dose-response curves (see section 0). The model demonstration in Schmitt et al. (2015) used fictive data on reproductive and developmental effects (reduced litter size, pup survival, and delayed maturation) that followed a linear dose-response relationship above NOAEL (see section 0). Data were derived from mean toxicological profiles across 15 commonly used pesticides.

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed for the model in general but only for a specific application in pesticide risk assessment. In the model demonstration by Bastiansen et al. (2013) no or only minimal pesticide effects on the vole population size were detected (see section 2.8.1.2). In Schmitt et al. (2015), the modelled pesticide concentrations imposed up to 28 % reduction in litter size and up to 2.4 d delay in maturation. No significant effects on the population size in every January were predicted unless the application rate was increased 10x or when the mortality caused by each application was set to 20 %. The authors concluded that the studies demonstrate how population models such as eVole can be successfully applied in risk assessment to show that minimal pesticide effects on the individual level do not translate into non-negligible effects at the population level. A critical discussion concerning the limitations of such models has not been provided in these studies.

2.8.3 Model Evaluation

2.8.3.1 Problem Definition

The regulatory context in which the model is run

For Higher Tier assessments on vertebrates, the actual protection goal of "no visible mortality and no long-term repercussions for abundance and diversity" has been defined (EFSA 2009b). In this context, eVole may be potentially applied to assess whether sublethal effects observed in toxicity tests will result in any adverse population effects. With simulation studies the risk assessment may be extended to a range of different scenarios, particularly for vertebrate species for which the number of experimental studies needs to be minimized for ethical reasons.

The question that has to be answered with the model

Some of the studies published to demonstrate the applicability of the model in risk assessment set a focus on the prediction of recovery times after PPP exposure. However, it should be noted that temporary effects on the abundance and population structure of mammals for days to weeks have been considered acceptable only if caused by avoidance behaviour (EFSA PPR 2010). Because eVole (currently) does not simulate avoidance-induced behavioural changes and its potential eco(toxico)logical consequences, the focus of eVole in risk assessment should be to detect the occurrence of any adverse population effects. This was done e. g. in Schmitt et al. (2015), though on a very coarse time scale (population sizes reported once per year).

The available knowledge and data relevant to the risk assessment question

The model is based on substantial knowledge and data on the behaviour and demography of common voles. However, eVole does not simulate the organic development or health of individuals, therefore sublethal effects observed in laboratory studies need to be translated to a decrease in survival or reproduction prior to modelling. Fitting the toxicity module for individual-level effects to toxicological data will thus require a set of assumptions on the specific mode of action and on how individual-level effects will act under realistic conditions outside the laboratory, where they may interact with additional stressors. These assumptions are crucial for the modelling outcome and an important source of uncertainty. They need to be thoroughly justified and communicated.

The outputs required to answer these questions including performance criteria for the regulatory model

The model provides various endpoints that may be suitable to address the actual protection goal mentioned above. For example, Schmitt et al. (2015) used the following endpoints as specific protection goals: (1) No long-term decline of population density during consecutive use of the product under evaluation, and (2) deviations between densities of the control and affected populations <5%.

The species to be modelled

The model was developed for small territorial mammals. The common vole and the common shrew have been listed as representatives of the generic focal species "small herbivorous mammals" and "small insectivorous mammals", respectively (EFSA 2009b). Due to high reproduction rates, particularly the common vole shows low vulnerability (Wang 2013) and is therefore principally not a suitable model species according to the EFSA Sci. Op. on GMP (EFSA PPR 2014b). However, in Central Europe both generic focal species cover no other real species that are considerably more vulnerable, therefore

it seems appropriate to base the risk assessment for small herbivorous and insectivorous mammals on the common vole and common shrew.

Requirements for the environmental scenarios to be used in the risk assessment

Setting up the environmental scenario in eVole requires information on the landscape composition and the seasonal variation in food availability for each habitat type. In the model, vegetation height is used as a proxy for the amount of food available. However, it must be noted that the link between vegetation height and food availability is specific for each habitat type and should be justified with appropriate field data. This is relevant to avoid an unrealistic distribution of voles in on- and off-crop area that may result in an underestimation of exposure. Information on the landscape composition and vegetation height may be available for specific areas, but so far, no agreement has been achieved on a realistic worst-case scenario for whole European member states. The dietary exposure in eVole is calculated according to the guidance for risk assessment on birds and mammals (EFSA 2009b) and therefore uses the information available from classical ERA. Since eVole simulations will be submitted as part of Higher Tier studies, it may be appropriate to consider also additional exposure routes such as dermal and respiratory exposure which is currently not supported.

2.8.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

The authors cite comprehensive data on the population dynamics and home range dynamics across the seasons in different habitats that appear appropriate for the development of the basic model without (pesticide effects). Ecotoxicological data used for individual-level PPP effects have to be evaluated case-by-case.

Assessing the risk of adverse population effects requires the consideration of additional stressors (e. g. species interactions, farming practices) that may increase the vulnerability in the field. Though not explicitly modelled, additional stressors have been covered implicitly to the extent they were experienced in the field studies used for parameterization. Some important attributes (e. g. duration of gestation and lactation, litter size) were parameterized using both field and laboratory studies. However, the given parameterization based on field data is only valid for Central Europe; modelling scenarios with different climatic conditions will require substantial re-parameterization.

The model includes also several ad-hoc assumptions for which no data exist to set up behavioural rules in the context of home range dynamics and dispersal. In particular, no observational data are available to justify the random walk algorithm for dispersal or the algorithm for home range optimization. E. g., the model assumes no avoidance behaviour that might result in the abandonment of contaminated home ranges due to decreased available food, and consequently in reduced reproduction.

Has the quality of the data used been considered and documented?

The quality of the data has not been discussed. However, the cited data have been published in peer-reviewed scientific articles, suggesting sufficient quality for the modelling purpose.

Have all available data been used? If not, is there a justification why this information has not been used?

The selection of data used for the modelling has not been justified. We are not aware of additional data that could have been used.

2.8.3.3 **Conceptual Model**

Are the specific protection goals sufficiently well addressed by the model?

eVole simulates all individuals of a vole (or shrew) population in daily time steps at a field- to small landscape scale with a resolution of 5 m, distinguishing three life stages with separate attributes (mortality, food demand, territorial behaviour). These entities and attributes are principally appropriate to address the protection goals of no long-term decline and no observable adverse population effects. However, the model is limited to the assessment of direct toxic effects of PPP. The duration of indirect effects that may result from avoidance behaviour - which are also addressed by the specific protection goal described in EFSA PPR (2010) - cannot be analysed with eVole. This would require an extended algorithm for home range selection that weighs avoidance vs. food limitation, and possibly the simulation of direct effects of food limitation on development and reproduction.

Are the modelling endpoints relevant to the specific protection goal?

The main model output is the development of population density over time along with confidence intervals that result from a Monte-Carlo approach with most parameter values drawn from random distributions. The results can be directly applied to various relevant protection goals (long-term decline, percent decrease compared to control, risk of extinction).

Is the modelling approach justified?

In contrast to other IBMs such as IDamP or the *Chaoborus* IBM population model, eVole is rather based on rules than on equations. Even more than in other models, population dynamics in eVole are therefore an emerging property that cannot be predicted from looking at the individual functions. This modelling approach is generally appropriate. However, it is unclear whether the additional complexity introduced with home ranges actually leads to improved predictions on adverse population effects of PPP, as compared to simpler modelling approaches that relate reproduction directly with population density. Similarly, it is unclear whether spatially explicit modelling actually increases ecological realism in model predictions over a simpler approach that may just draw a proportion of feeding on- and off-crop for each individual from a (seasonally varying) probability distribution. The model uses reflecting boundary conditions (voles cannot leave or enter the grid borders) so that edge effects may be possible. Additionally, interactions of PPP effects and additional stressors on the individual level are not simulated which provides a relevant source of uncertainty in the results. Finally, the ecological realism of modelling a density-regulated population without the community context that may have a strong impact on its recovery potential may be questioned.

Is the conceptual model logical?

The conceptual approach is principally logical. However, potential inconsistencies have been identified regarding the implemented mechanisms of density regulation. In the model, the individual development and survival does not dynamically interact with food availability or population density, so that density-regulation proceeds only in a delayed way through decreased reproduction. Food limitation can therefore affect population size only during the breeding season, so that the model is limited to relatively stable environmental conditions. E. g., in eVole a large amount of food may be removed in autumn due to seasonal die-back or farming practice. Real-world populations would probably decrease in winter due to starvation and need to recover through increased reproduction in spring when food grows again and the next breeding season starts. PPP effects on reproduction may considerably delay population recovery from these additional stressors in spring. In contrast, in eVole, population size would not be affected from the simulated food limitation outside the breeding season. In spring,

population size would not need to recover from this stress and be therefore less sensitive to PPP effects on reproduction that may be masked by density regulation. However, populations in eVole generally need to recover in the breeding season from background mortality in winter. With increasing strength of this seasonal population cycle (depending on the environmental scenario), reproductive effects of PPP in spring may be similarly observed with and without food limitation in the model.

Individuals in eVole can maintain home ranges with insufficient food resources, down to 40% of their requirement. It seems not logical that such a reduction in resources does not directly impact reproduction or survival. However, individuals in eVole will try to increase their home ranges following food limitation. This decreases the number of new home ranges that can be established and thus indirectly affects reproduction (and optionally survival) in the next generation.

The model description suggests that a landscape cell provides its full food value to each of the individuals that share this cell as part of their home ranges. Since only adult females expel each other in the common vole model, this means that all other individuals do not suffer from competition (an indefinite number of subadults could share the same home range). This facilitates fast recovery of the population size in the model, although in real populations, territoriality of established survivors and competition among subadults may limit the growth of the young generation.

Are the processes included in the model relevant to the addressed issue?

eVole includes relevant processes such as the temporal variation in the spatial distribution of individuals between on- and off-crop areas, density-regulation and seasonal variation in the population structure. However, it is not clear to which extent home range dynamics need to be modelled explicitly in order to assess how PPP affect the population size and structure (see above). In contrast, some relevant processes that may increase the sensitivity of vole or shrew populations to PPP have not been considered. The model assumes that the individual sensitivity to PPP does not change with environmental conditions. However, populations have shown to be particularly sensitive to sublethal effects of PPP when they are close to carrying capacity, because crowding stress may increase the susceptibility of individuals to PPP effects (Knillmann et al. 2012a). Additionally, abiotic stressors may increase the sensitivity as compared to those observed under standard laboratory conditions, though other aspects such as avoidance may also decrease PPP effects. Finally, the potential for population recovery may considerably decrease when the individuals have to compete with other species (Knillmann et al. 2012b) or when the population is controlled by predators such as birds of prey.

Are the links between different processes to the variables logical?

The established links between the different processes are logical. However, see the missing links of population density and the environmental context with the individual development and sensitivity to PPP effects outlined above.

Are the temporal and spatial scales relevant in regard to the problem definition?

eVole uses an appropriate temporal and spatial scale for the population modelling. Regarding exposure, the time step of 1 d can be problematic for fast dissipating pesticides and may require a maximum or time-weighted average concentration to be used in the sub-model for individual-level effects.

2.8.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

The algorithms related to demographic processes are quite trivial and do not seem to require specific justifications. The algorithm that determines home range dynamics is not explicitly justified. The rules

for intraspecific competition were established from field observation on the spatial distribution of the different life stages. However, the implicit assumption that growth and survival are not affected by the existence and quality of a home range is not justified by the authors. Dispersal based on a random walk is not explicitly justified, but it is quite obvious that this option was chosen because alternatives (e.g. correlated random walk) require data for parameterization that is not available in the literature.

Are the most important mathematical equations described?

The basic population model is rule-based and does not include any equations. The equations for the PPP exposure and effects modules have been described comprehensively in (RIFCON 2018).

Is there a description of the variables and parameters including their meaning and unit?

The variables and parameters together with their units have been described appropriately (Wang and Grimm 2007, Wang 2013, RIFCON 2018).

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

A verbal justification is provided specifically for the modelling of home range dynamics, which is the most complex part of the model. The main argument here is that in territorial species, home range dynamics are an important mechanistic determinant of population dynamics. Unfortunately, no analyses have been conducted to demonstrate that the explicit modelling of home range dynamics is necessary for the reliable assessment of adverse population effects of PPP.

Are references supporting the equations been provided?

The basic population model does not include any equations. The rules of the model have been established based on life cycle information published in scientific literature that has been cited in the model descriptions. Equations for the calculation of a DDD in the exposure module were taken from the EFSA guidance on risk assessment for birds and mammals (EFSA 2009b). Equations for the modelling of individual-level toxicological effects have not been referenced but include standard procedures used in ERA such as log-logistic dose-response curves. Equations used to convert observed effects on body weight to effects on survival were taken from a peer-reviewed publication (Oksanen et al. 2007).

2.8.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

Flow charts, biological parameter values and descriptions of the most relevant algorithms incl. examples with pseudo-code have been provided in (RIFCON 2018). A technical description incl. all algorithms and numerical methods is not publicly available.

Is the computer code well readable and is it available?

The computer code is not publicly available. It can be provided to authorities upon request, for evaluation.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

According to Wang (2013), the model for the common vole has been subjected to “a thorough revision of the code and a manual inspection of variables during test runs. Subroutines and sub-models were

tested separately. Additionally, visual debugging was applied to test subroutines and sub-models, for example a visual inspection of home ranges or landscape parameters." In (RIFCON 2018), extensive verification of implementation has been presented for each of the modules.

2.8.3.6 Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

The default biological parameterization was based on field studies on Central European shrew and vole populations in various habitats that represented realistic, but potentially not worst-case scenarios in terms of field size, farming practice etc. eVole does not provide a default scenario for the landscape composition and landscape events, because these settings are case-specific. Studies for the demonstration of model applicability used various settings that represented realistic (but not worst-case) landscapes with different patch sizes and hedge elements. Schmitt et al. (2015) state that further investigation and agreement are needed to develop recommendations for landscape attributes such as size, structure, and crop rotation to define appropriate regulatory risk assessment scenarios.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

Biological parameters have been inferred from cited literature. In the scenarios used for model demonstration, food availability and vegetation cover have been inferred from cited literature or set to pre-defined values in order to analyse the model behaviour.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

eVole models the dietary exposure from foliar PPP residues on treated vegetation. Initial foliar residues of contaminated food after PPP application need to be imposed on the model. eVole then applies single first-order kinetics to calculate the daily dissipation, and the DDD is calculated based on the case-specific proportion of contaminated food (set by the user). However, exposure from inhalation or dermal contact during spraying is not modelled.

Is the level of conservatism placed into the scenarios appropriate?

This item cannot be addressed for the model in general. It can by only addressed to a specific application in the context of environmental risk assessment.

2.8.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

Calibration of the food demand of voles relating vegetation height to home range size was adequately described in (RIFCON 2018). Calibrations for the version on the common shrew are not sufficiently described.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Parameter values were derived from peer-reviewed scientific studies, and the ability to reproduce population dynamics observed in the field (see section 2.8.2.9) suggests that parameterization was successful.

Were the estimated parameter values realistic?

There is no indication that the estimated values were unrealistic.

Are the data sources sufficiently documented?

For parameter values derived from the literature, data sources were sufficiently documented.

2.8.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

The sensitivity analysis has been adequately documented.

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

A local sensitivity analysis was performed for the basic model without pesticides. In a narrow sense, this analysis is valid only for the tested environmental scenario (mixed landscape of grassland, pasture and arable fields). However, results show that without pesticide exposure, model predictions on the population density, the production of offspring and population growth are relatively insensitive to changes in the parameter values studied; it can be expected that the model behaviour will not fundamentally change with other realistic landscape compositions. However, no sensitivity analysis has been performed for the model behaviour in the presence of PPP effects. E. g., it would be important to identify the minimum size of imposed individual-level effects that will lead to observable adverse population effects in typical settings.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

The results of the sensitivity analysis have been presented in comprehensive tables in (RIFCON 2018).

Has the uncertainty analysis been adequately documented?

No uncertainty analysis has been performed. An uncertainty analysis is most useful for a specific model application and should be performed case-by-case.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

No uncertainty analysis has been performed.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

No uncertainty analysis has been performed.

Uncertainty is propagated to the model results?

The probabilistic approach of the model covers for the expected uncertainty in input parameters that are provided as probability distributions. This uncertainty is propagated as confidence intervals to the model results. Other sources of uncertainty (e. g. structural) cannot be propagated.

Have confidence intervals been estimated and has this information been used in further model use?

See above.

2.8.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

No performance criteria have been predefined before model validation. The aim was probably to reproduce the observed patterns in population dynamics as close as possible.

Are the model outputs that are compared relevant in view of the problem definition?

The predicted and observed dynamics of population sizes and structure have been compared to validate the basic population model without pesticide exposure. However, model predictions on adverse population effects of pesticides and on population recovery from such effects have not been tested with data from experiments or field studies. Therefore, the ability to predict population level-effects of natural populations from Tier 1 data has not been validated.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

Data for validation was obtained from peer-reviewed studies, but these studies were not designed for the purpose of model validation.

Is the dataset relevant in view of the problem definition?

The data used for validation were partly relevant (see above). While it has been demonstrated that several patterns in the population dynamics of natural populations without pesticide exposure can be reproduced in eVole, no data-based comparison has been performed for the main problem of how individual-level pesticide translate into population effects in common voles.

Is the fit of model output to the data good enough?

The authors concluded that the fit was sufficient, though this is difficult to assess without defined performance criteria. Altogether, the model was generally able to reproduce the investigated spatial and temporal aspects of the population and home range dynamics reasonably well. However, in contrast to the assumptions in eVole, the field data suggest that populations produce a low number of offspring also outside the breeding season (Fig. 8). Assuming no reproduction outside the breeding season may be conservative when modelling pesticide exposure during the breeding season, but not conservative when modelling exposure outside of the breeding season.

Has the performance of the model been reported in an objective and reproducible way?

Results of the studies for model validation have been presented in an objective and reproducible way (Wang 2013, RIFCON 2018).

2.8.3.10 **Model Use**

Is a user manual available?

A user manual is not publicly available but can be provided to authorities as part of a specific dossier, together with code and model executable.

Have all aspects of the modelling cycle been documented?

The model is well described based on the ODD protocol (Grimm et al. 2006) in Wang and Grimm (2007) and in Wang (2013), and based on the TRACE protocol (Grimm et al. 2014) in RIFCON (2018).

Has a summary sheet been provided by the modeller?

No summary sheet as recommended in the Sci. Op. on GMP (EFSA PPR 2014b) has been provided for the general presentation of the model to the public. It is submitted as part of the dossier for projects conducted after the issuing of the EFSA scientific opinion, in addition to TRACE and a completed evaluation checklist.

2.8.3.11 **Suitability of the Model for Regulatory Purposes**

Is there a possibility for dialogue between the modeller and the risk assessor?

Model developers are available for dialogue with risk assessors at info@rifcon.de. The model has been published in scientific papers (e. g., Wang and Grimm 2007, Wang 2013) and presented at conferences (Bastiansen et al. 2013, Ibrahim et al. 2019), and the TRACE document is available to risk assessors up on request**Fehler! Linkreferenz ungültig.**.

Is a version control system implemented?

Model development is conducted under version control (RIFCON 2018).

2.8.3.12 **Overall Judgement**

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

eVole has been designed to investigate whether pesticide effects observed on individuals will translate to adverse population effects in the field. The model is intended to support and complement Higher Tier studies on mammals (such as dietary experiments and field studies) for which experimental data are restricted for ethical and financial reasons. The model species (common vole and common shrew) are characterized by low ecological vulnerability, but are representative surrogate species for the guilds of small herbivorous or insectivorous mammals to be protected.

For small mammals, the actual protection goal comprises no visible mortality and no long-term repercussions for abundance and diversity of non-target species (EFSA 2009b). eVole predicts the population size and structure in a modelled landscape over time and therefore allows to address this actual protection goal. However, further analyses should be done to assess the power of the model to predict population effects, i. e. to identify minimum individual-level effect sizes that will generate adverse population effects. Temporary population effects have been considered acceptable if caused indirectly by avoidance behaviour. Because eVole does not model avoidance behaviour and its potential ecological

consequences (starvation, increased susceptibility to predation), this specific protection goal cannot be properly addressed.

While eVole addresses population dynamics on appropriate spatiotemporal scales, the model might be overly complex for its purpose. The explicit simulation of home range dynamics certainly adds realism to the model but also adds substantial complexity and requires a high spatial resolution. It is unclear whether and to which extent this complexity improves predictions on adverse population effects, while might substantially slow down computations on a large spatial scale. Additionally, potential inconsistencies in the mechanisms of density regulation have been identified: In eVole, individuals can maintain a home range that provides only 40 % of their required food without suffering from any negative effects. Additionally, it seems that if space is limited, many individuals (except adult females) may share the same territory without facing negative effects from competition. Because individuals do not starve to death or expel subadults, the population can quickly recover from pesticide effects.

In contrast, a number of relevant ecological mechanisms that potentially affect the effects of pesticides are not directly addressed in the model: Pesticide exposure is assumed not to affect the spatial behaviour or the efficiency in food utilization. However, effects on food utilization may be indirectly included e. g. via a modelled pesticide effect on body weight as consequence of decreased food utilization (but may not aggravate with food limitation). Effects of biotic and abiotic stressors on the recovery potential are partly included implicitly, because the background mortality was parameterized with field studies where individuals were exposed to natural environmental stress. Other traits such as the duration of gestation and lactation, and litter size were partly parameterized using laboratory studies without natural stressors but, according to the modellers, in agreement with the available field studies from the literature (personal communication). Additionally, individual-level effects of pesticides are assumed not to increase with environmental stressors such as food scarcity, density stress or the risk of being preyed. These implicit assumptions have not been justified and are important sources of uncertainty in the model predictions. eVole reproduced patterns of population dynamics in field studies without pesticide exposure. However, before the model may be applied in ERA its ability to predict adverse population effects of pesticides should be validated.

The toxicity module in version 3.0 offers to model individual-level pesticide effects using static dose-response curves, whereas the dietary exposure in eVole varies every day as a result of the modelled dissipation and home range dynamics. The module thus allows to link experimentally observed effects to one of several metrics of exposure (DDDt, TWA21, Max21, see model description above). Each of these options is associated with different implicit assumptions that are more or less conservative. E. g., fitting daily mortality in the model to 24 h acute toxicity data will assume that effects neither build up over prolonged time nor that individuals will develop tolerance. Linking chronic effects to the MAX21 assumes that effects are driven by short-term peak exposure, which is generally more conservative than the TWA21 approach. However, applying the MAX21 might be too conservative for effects on body weight which are translated into daily mortality, because the maximum mortality will be executed each day even 20 days after the peak exposure has gone. Hence, individual-level effects in eVole should be implemented with great care. The built-in toxicity module was designed to handle minimalistic data. However, it seems preferable to link eVole to more sophisticated modules designed to handle-time variable exposure (e. g. GUTS, DEBtox) when sufficient data are available to fit these modules.

Altogether, the model seems principally suitable to investigate the effects of various landscape compositions and farming practices (incl. pesticide application patterns) on the risk of pesticides for small mammals. These results may be used to support risk management. However, the selection of implemented mechanisms appears somewhat biased towards processes that potentially lead to an underestimation of the risks of pesticides (see below). Due to these uncertainties in model predictions, it is difficult to derive an absolute level of risk for real populations in the field from the modelling results (i. e. an accurate probability for a given percentage in population decline or an accurate time to population recovery). Conclusions on the risk for populations in the field should be drawn with care.

2.8.4 Qualitative Assessment of Uncertainties

2.8.4.1 Potential for Underestimation of Real Risk

- Additional real risk may arise from toxicological effects that are not considered in the model, e.g. effects on body weight that translate later into effects on population dynamics or changes in spatial or competitive behaviour or decreased food utilization efficiency leading to increased food demand.
- Individuals do not suffer from reduced food availability down to 40 % of their food demand and do not suffer from intraspecific competition (except for adult females). As a consequence, the age structure (and population size) of a population will quickly recover after pesticide exposure because the development of the young generation is not hindered by increased competition with individuals of the same age.
- Some life history traits that are highly relevant for population recovery (duration of gestation and lactation) were parameterized using laboratory studies with artificial conditions in terms of low biotic or abiotic stress. The recovery potential of populations in the field may thus be overestimated, though gestation and lactation was in agreement with field observations according to the modellers.
- Biotic and abiotic stressors that are not present in toxicity tests may increase the individual-level effects of pesticides.
- When applying the model to the insectivorous common shrew, the toxicity module for individual-level effects needs to consider potential biomagnification along with the food chain.
- The only exposure pathway modelled in the included toxicity module is the dietary exposure. Dermal or inhalation exposure during spraying is not considered.

2.8.4.2 Potential for Overestimation of Real Risk

- The model proceeds in time steps of 1 day. For fast acting and fast dissipating pesticides, this temporal scale might be too coarse and pesticide effects calculated on the basis of predicted peak exposures might be overestimated.

2.8.4.3 Potential for Uncertainty in Either Direction

- Pesticides may affect the spatial behaviour of individuals (avoidance). Avoidance may reduce exposure and thus direct effects of pesticides, but may lead to indirect effects (e. g. abandonment of home ranges, followed by decreased reproduction).
- An important source of uncertainty is the apparently open unsettled question of which landscape scenarios (field size, arrangement of fields, additional farming practices) are most suitable for the intended risk assessment.
- In eVole 3.0 for the common vole, the toxicity module for individual-level effects uses static dose-response curves on a daily basis to depict daily effects and the exposure of individuals in the model varies every day. The dose-response relationships can be fitted either to acute or chronic data, and effects can be then calculated using either the daily dietary dose, the time-weighted dietary dose over 21 d, or the maximum dietary dose of the last 21 d. The different approaches are associated with varying levels of conservatism and need to be selected case by case.

2.9 ALMaSS and Wood Mouse Model: IBM Population Models for Mammals

Evaluation by Stephanie Kramer-Schadt

2.9.1 General Information

2.9.1.1 Background and Concept

ALMaSS

ALMaSS (Topping et al. 2003) is a flexible simulation platform integrating animal population dynamics under spatial management scenarios (resolution 1 m^2) in a spatially-explicit individual-based modelling (IBM) framework. The external drivers of the system, that is land management and resource/ vegetation growth together with climatic effects, are described in the landscape model, and the reaction of the respective animal in the individual-based animal model (IBM).

A detailed overview over the basic functionalities is given in Topping et al. (2003)[17], compiled in the following: The landscape model comprises detailed mapping in polygon and raster format at a minimum of 1 m^2 scale, weather, agricultural management, and vegetation growth. Weather data is stored as daily records of mean temperature, mean windspeed, and total daily precipitation and interfere with vegetation growth. There is a basic division between vegetated and non-vegetated landscape polygons. Vegetated polygons have their own models to describe the growth of vegetation during the year on a daily basis. In the case of farmed areas, management is modelled in detail and consists of crop rotation, such as spring barley, winter rye and maize, and crop husbandry. Each crop type has its own crop husbandry plan, which consists of a series of timed events, dependencies and conditions. The events, such as harvest, are recorded when they take place in a field polygon and this data can then be accessed by the animal models. An event may also cause other changes in the simulation, for example, harvest will alter the vegetation biomass of a crop, and herbicide applications will affect the weed biomass.

Animal models are agent-based, designed using the state/transition concept (see below), and are rule-based, that means 'if-then' conditions describe the behaviour of the individuals. Each animal may interact with other animals and directly with its local environment. Simulations of crop diversity and rotation lead to significant effects of spatial and temporal heterogeneity on population sizes, population fluctuations and landscape permeability. Spatial and temporal heterogeneity interact and thus different responses to temporal factors occur at different levels of spatial heterogeneity. ALMaSS was originally designed for policy questions regarding the effect of changing landscape structure or management on key animal species in the Danish landscape (e.g. removal of hedgerows, effects of organic farming,…) using indicator species such as insects, birds and mammals (see Table 10 from Jepsen et al. (2005) below).

Table 10: ALMaSS – Entities and Processes in Different Species Models

Species model	Behavioural states
Carabid beetle (*B. lampross*)	Initiation (E, L, P, f), development (E, L, P), hatching (E, L), larval dispersal (L), emerging (P), reproduction (f), dispersal (f), aggregating (f), hibernation (f), dying (E, L, P, f)

[17] Topping, C. J., T. S. Hansen, T. S. Jensen, J. U. Jepsen, F. Nikolajsen and P. Odderskaer (2003): ALMaSS, an agent-based model for animals in temperate European landscapes. Ecological Modelling 167: 65-82.

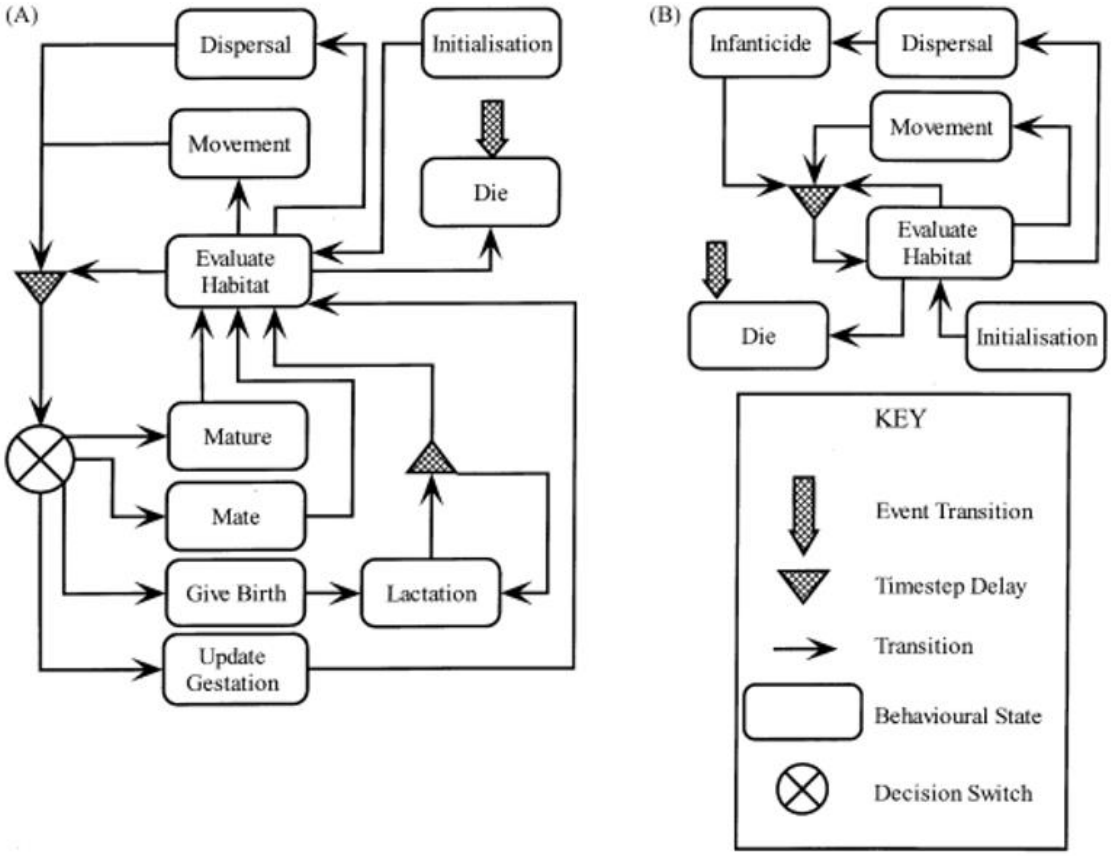
Species model	Behavioural states
Linyphiid spider (*O. fuscus*)	Initiation (E, J, f), development (E, J), hatching €, maturation (J), dispersal (J, f), assess food (J, f), reproduction (f), dying (E, J, f)
Skylark (*A. arvensis*)	Initiation (C, N, P, m, f), development (C, N, P), hatching (C), nest leaving (N), maturation (P), migration (m, f), flocking (m, f), find territory (m, f), establish territory (m), give up territory (m, f), build-up resources (f), start new brood (f), attract mate (m, f), mate guarding (m), make nest (f), lay eggs (f), incubation (f), parental care (m, f)dying (C, N, P, m, f)
Field vole (*M. agrestis*)	Initiation (m, f), evaluate and explore (m, f), habitat assessment (m, f), dispersal (m, f), maturation (m, f), lactation (m, f), giving birth (f), mating (f), infanticide (m), dying (m, f)
Roe deer (*C. capreolus*)	Initiation (J, m, f), maturation (J), habitat assessment (m, f), feeding (m, f), ruminating (m, f), dispersal (m, f), establish range (m, f), establish territory (m), social ranking (m), mating (m, f), giving birth (f), winter grouping (f), dying (J, m, f)

Object types: Egg (E), larva (L), pupa (P), adult female (f), adult male (m), juvenile (J), clutch (C), nestling (N), prefledgeling (P). Data from Jepsen et al. (2005, edited).

All animal models are behaviour-based and built upon a state/transition principle. Thus, an animal is considered to be in a specific state when it exhibits specific behaviour. When a condition or conditions are fulfilled, there is a transition to another state. Transition conditions may be probabilities, or internal or external events (e.g. giving birth or being eaten). Using this method, large and complicated sets of behavioural rules can be built up from a set of simple building blocks (Topping et al. 2003); an example of a process overview for the field vole is shown in Figure 16 below). The main processes growth, reproduction, mortality as well as dispersal and territoriality are based also on habitat quality; density-dependence hence is an emerging property based on habitat quality and territoriality. Each of the main processes has many sub-processes, e.g. for the species 'vole', reproduction consists of maturation, mating, giving birth and lactation. Model parameters (see e.g. Table 10, above) are based on literature data.

To be clear, ALMaSS was initially not set up for risk assessment of chemical substances, but for a population viability analysis under different spatial configurations of agricultural fields and land management regimes, like crop rotation and harvesting with removal of suitable shelter habitat or food resources. Since ALMaSS is based on the animal's ability to move between (treated) patches, it was then extended to simulate the take-up of different doses of a pesticide. I.e., with ALMaSS, the effects of different spatial applications of a pesticide in relation to habitat configuration and species' behavioural processes can be assessed at the population and landscape level. Individual-level endpoints derived from toxicological studies, e.g., body weight of parents and offspring, litter size, viability of pups, are therefore translated into population level effects (Schmitt et al. 2015). This is done for some species with toxicokinetic (TK)-like calculations, as in the European brown hare (*Lepus europaeus*) model application (Topping et al. 2016). ALMaSS examples used in environmental risk assessment (ERA) follow the EFSA Sci. Op. on GMP (2014b) and are therefore well documented and follow the guidelines at their best.

Figure 16: ALMaSS – Simulation Framework of the Vole Model



State/transition diagram for female (A) and male (B) voles (*Microtus arvestis*) showing the behavioural states modelled and their relationships. Graph reproduced from Topping et al. (2003).

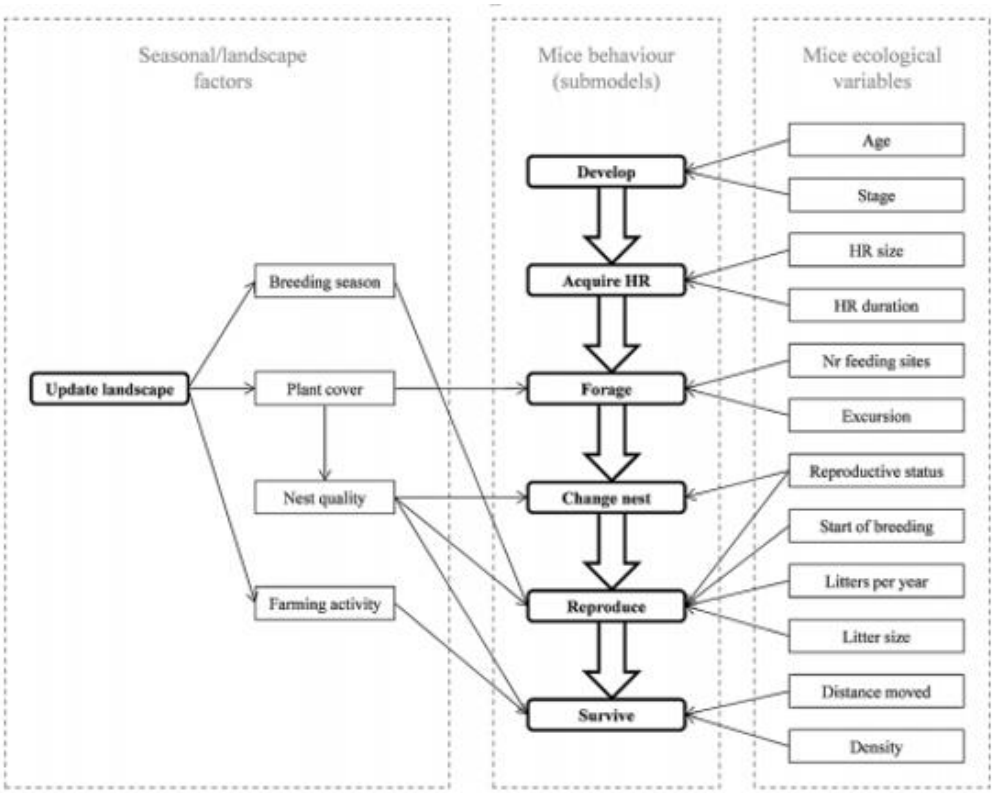
Wood Mouse Model (Liu et al.)

The wood mouse (*Apodemus sylvaticus*) model (Liu et al. 2013, Liu et al. 2014; see also comparison in Schmitt et al. 2015)[18,19] follows the same principles as ALMaSS regarding documentation in scientific journals and ODD (Overview, Design concepts, and Details) documentation (Grimm et al. 2006), the standardized description of individual-based and agent-based models. Also, the design of the individual-based model follows basically the same processes at the individual level as the vole model, shows the same drawbacks regarding population regulation process in a community context (see below) and includes a TK-part similar to the brown hare model in ALMaSS. The following figure 17 from Liu et al. (2013) shows the similarity of this model with ALMaSS in the basic processes of development/maturing, moving and acquiring a home range, foraging, survival and reproduction:

[18] Liu, C., R. M. Sibly, V. Grimm and P. Thorbek (2013): Linking pesticide exposure and spatial dynamics: An individual-based model of wood mouse (Apodemus sylvaticus) populations in agricultural landscapes. Ecological Modelling 248: 92-102.

[19] Liu, C., A. J. Bednarska, R. M. Sibly, R. C. Murfitt, P. Edwards and P. Thorbek (2014): Incorporating toxicokinetics into an individual-based model for more realistic pesticide exposure estimates: A case study of the wood mouse. Ecological Modelling 280: 30-39.

Figure 17: Wood Mouse Model – Flowchart of the Conceptual Model



Overview of the wood mouse (*Apodemus sylvaticus*) model by Liu et al. (2013, 2014). Processes (bold boxes) and key variables (normal boxes) are indicated. Fat arrows indicate the order of processes and thin arrows influences. HR: home range, Nr: number. Graph reproduced from Liu et al. (2013).

An example of the number of parameters and variables needed for the wood mouse IBM processes is shown in the following table from Liu et al. (2013). Most of the biological parameters are based on peer-reviewed literature.

Table 11: Wood Mouse Model – Key Parameters, Values and References

Category	Name / meaning (notes)	Unit	Range / defaut value	Reference
Home range and feeding	Habitat quality (=plant cover)[a]		Good (when plant height ≥ 30 cm and cover ≥ 50 %) Bad (when plant height < 30 cm or cover < 50 %)	Ouin et al. (2000)
	Home range sizes	m^2	1424	Wolton (1985)
	Home range duration[b]	days	[10, 30]	Wolton (1985)
	Nr. feeding sites (daily)[c]		10 (breeding season) [3, 5] (non-breeding season)	Jealott's Hill (un-published)
	Max. nr. of mice in one nest		1 (breeding season) 3 (non-breeding season)	Wolton (1985)
	Excursion distance	M	> 25 m	Wolton (1985)
	Daily excursion probability		2 % (adults breeding season) 4 % (juveniles breeding sea-son)	Wolton (1985)

Category	Name / meaning (notes)	Unit	Range / defaut value	Reference
			5 % (adults non-breeding season) 10 % (juveniles non-breeding season)	
Reproduction	Breeding season[d]		[March, November]	Flowerdew and Tattersall (2008)
	Date of first reproduction[e]	day	[01. March, 30. April]	Tattersall and Macdonald (2003)
	Wean age	day	20	Flowerdew and Tattersall (2008)
	Mature age	day	60	Macdonald and Barrett (2005)
	Max. litters per age[f]		6	Macdonald and Barrett (2005)
	Litter size[g]		Base line 2 or 3, in June and July: +1, in August: +2	Smyth (1966), Macdonald and Barrett (2005), Flowerdew and Tattersall (2008)
	Pregnancy duration	days	25	Ashby (1967)
	Lactation duration	days	20	
Survival	Age	days	[1,600]	Macdonald and Barrett (2005)
	$d = \sum c_1$: daily mortality rate: c_1: increased mortality caused by movements in nest changing[h]		0.001 * distance	Calibrated
	c_2: increased mortality caused by excursion		0.05	Calibrated
	c_3: global density-dependent mortality[i]		When pop. Size > 1600: 0.0006 per surplus mouse i. e. 0.0006 * (pop. Size – 1600)	
	C_4: increased mortality caused by poor nest[j]		0.1	
	c_5: local density-dependent mortality for mice in different life stages[k]		During [February, May]: 0.005 * local mice pop size (for infants and juveniles) 0.005 * local adult pop size (for adults) During other months: No density-dependent mortality	Macdonald and Barrett (2005)

Category	Name / meaning (notes)	Unit	Range / defaut value	Reference
	c_6: increased mortality caused by farming practices		0.04 (plough, dig) 0.02 (harrow, sow, cultivate, plant) 0.1 (harvest)	Calibrated Calibrated Macdonald and Barrett (2005)

Notes:
[a] In this model, food abundance is not considered.
[b] Time duration of how long a home range remains the same.
[c] Dependent on active hours in different season.
[d] Typical breeding season is March-October, but sometimes can continue throughout winter.
[e] Breeding females start to appear in February / March and all females are in reproductive status before May. Onset of breeding varies between years; therefore, it is assumed there is no clear pattern for the time of first reproduction, thus a uniform distribution within the reported range of dates.
[f] Number of litters a mouse can produce during one year. Max of six successive pregnancies recorded in wild.
[g] Mean size 4-7, halved as only females are modelled. Litter size peaks in June-August. Winter litters are usually small.
[h] When a mouse moves longer than 50 m, otherwise $c_1 = 0$.
[i] See reference, section 2.2.3 for justification.
[j] Assume same as harvest effect.
[k] Density dependence is strong during population increase but is not acting during decrease. Spatial density-dependent female reproductive activity and territoriality limits the peak numbers. Local: within 50m radius; infants and juveniles are not competent enough to affect adult survival, so adults only count other adults as density whilst infants and juveniles have to count all stage groups.
Data from Liu et al. (2013, edited).

Further, the following variables and parameters are needed to link the species' biological processes to the treatment model via a TK-part.

Table 12: Wood Mouse Model – Parameters for the TK Part

Category	Name, meaning and calculation	Unit	Range / defaut value	Reference
Energy content and requirement	Seed energy content = FE * (1 – MC/100) * AE	kJ/g seeds	13.18	EFSA (2009b)
	FE (food energy)	kJ/g dry	18.4	
	MC (moisture content)		14.7	
	AE (assimilation efficiency)		0.84	
	Body weight (bw) when age ≥ 60 d): $bw = (1 – (1 – (m_0/m_\infty) * (1/3)))*e(-bt/3))3$ when age < 60 d: $bw = (3/t_w) * \ln((1 – (m_0/m_\infty) * (1/3)) / (1 – (mw/m_\infty) * (1/3)))$	g	21.7	EFSA (2009b), Sibly et al. (2013)
	tw (age at weaning)	Days	21	Corbet and Harris (1996)
	m_0 (body weight of neonates)	g	2	
	m_w (body weight of weanlings)		15	

Category	Name, meaning and calculation	Unit	Range / defaut value	Reference
	m_∞ (body weight of adults)		21.7	EFSA (2009b)
	Daily energy expenditure (DEE): DEE = $\log_{10}(a) + b * \log_{10}(bw)$	kJ		
	log10(a)		0.814	
	b		0.715	
Habitat and foraging	Probability for visiting newly drilled field 10h 5h 2h6h2h Intermediate	per visit	 0.048 0.094 0.116 0.5	Calibrated Assumed
	Daily excursion probability Adults Juveniles	%	 20 40	Wolton (1985)
	Winter wheat farming time Cultivate Harrow Sow Harvest	date	 Sep-25 Oct-04 Oct-05 Aug-25	
	Crop intervention (level of cover) BBCH[a] < 10 BBCH 10 – 19 BBCH 20 – 39 BBCH 40 – 89 Stubble	%	 0 (no cover) 25 (min cover) 50 (interm. cover) 70 (full cover) 20 (after harvest)	FOCUS (2001)
TK	$A_{1/2}$ (absorption half-life)	mins	10	
	<=> k_a (absorption rate= = ln(2/$A_{1/2}$)	per min	0.069	
	$t_{1/2}$ (elimination half-life)	mins	120; 60; 30	
	<=> k_a (absorption rate= = ln(2/$t_{1/2}$)	per min	0.0058; 0.012; 0.023	
Treatment and endpoint	Concentration on seeds (seed loading): Winter wheat seeds	Mg a.i. / Kg seeds	500	
	Dose-response curve: mortality rate = normsdist($\log_{10}(D_{eaten}$ or $C_{int} * \beta * \alpha$)			
	β (slope) for external dose (D_{eaten})		4.808	
	β (slope) for internal concentration (C_{nt})		4.808 ($t_{1/2}$ = 120) 4.804 ($t_{1/2}$ = 60) 4.800 ($t_{1/2}$ = 30)	
	α (intercept) for D_{eaten}		-9.615	
	α (intercept) for C_{int}		-9.161 ($t_{1/2}$ = 120) -8.886 ($t_{1/2}$ = 60) -8.491 ($t_{1/2}$ = 30)	

[a] Standard agronomic code to represent plant growth stages (Hess et al. 1997).
Data from Liu et al. (2014, edited).

Because these IBMs follow the same design principles, I basically focus on an example from the ALMaSS platform to evaluate the usefulness of the presented individual-based models for ERA. In the following, I will focus on the field vole model (*Microtus agrestis*) for the assessment as described in (Schmitt et al. 2015) and the brown hare model (Topping et al. 2016) if not stated otherwise. However, the highlighted problems that I will discuss with the vole model (e.g. lack of Tier 1 studies, lack of data and processes on top-down population control mechanisms like predation and competition) also apply for other species and taxa implemented in ALMaSS (e.g. birds) or in the wood mouse model (Liu et al. 2013, Liu et al. 2014).

2.9.1.2 **Status of the Model**

Currently, the ALMaSS platform has been used in >15 publications regarding applications to different species, rigorous testing of the IBMs developed, applications to ERA and (few) validations. Regarding ERA for pesticides, demonstrative applications of the model have been published for skylarks (Topping and Odderskaer 2004, Topping et al. 2005), beetles and spiders (Topping et al. 2014, Topping et al. 2015), voles (Dalkvist et al. 2009) and brown hare (Topping et al. 2016). ALMaSS comes along with a very good pseudo-code documentation of processes and parameters following the ODD (Grimm et al. 2006) that can be found at:

http://projects.au.dk/casesm/almass/

http://www2.dmu.dk/ALMaSS/ODDox/ALMaSS_ODDox/V1_01/main.html

2.9.2 Model Description

2.9.2.1 Problem Definition

Context in which the model will be used

ALMaSS is a platform in which models of case studies for predicting long-term effects of a pesticide to populations of small mammals or invertebrates in a spatial context are implemented. The wood mouse model by Liu et al. (2013) and Liu et al. (2014) is a stand-alone model based on the same modelling principles. This section is focussed specifically on the use of Individual-based population models for small mammals implemented in ALMaSS and the wood mouse model.

Specification of the question(s) that should be answered with the model

To develop recommendations for landscape attributes such as field size, structure, and crop rotation to define appropriate regulatory risk assessment scenarios.

Specification of necessary model outputs and protection goals

For terrestrial vertebrates exposed to pesticides, the specified protection goal regarding survival is no additional mortality; regarding chronic effects it is no effects on reproduction and therefore no long-term impact on populations. At higher tiers, effect assessments may also be directed to clearly establishing that there will be no long-term repercussions for abundance and diversity in the field. This is consistent with the legislation in place (Regulation (EC) No 1107/2009), as reported in the EFSA Sci. Op. on SPG (EFSA PPR 2010), where the goals of "no decline in biodiversity" and "negligible effects on population structure" are defined at the population level (Schmitt et al. 2015). The model outputs are therefore population development of the species under concern under various treatments/ scenarios.

Domain of applicability of the model

ALMaSS was set up for population risk assessment based on the animal's ability to move between (treated) patches and thus to take up different doses of a pesticide. I.e., with ALMaSS-IBMs, the effects of different spatial applications of a pesticide in relation to habitat configuration and species' behavioural processes can be assessed at the population and landscape levels. Individual-level endpoints derived from toxicological studies, e.g., body weight of parents and offspring, litter size, viability of pups, are therefore translated into population level effects (Schmitt et al. 2015).

Why is the model being used?

In cases where a high margin of safety can usually not be demonstrated using the standard approaches of deterministic risk assessment, individual-level endpoints derived from toxicological studies (e.g., body weight of parents and offspring, litter size, viability of pups) can be translated into population level effects in heterogeneous landscapes to indicate compliance with the protection goals. This concerns especially small herbivorous mammals (e.g., voles), since they often are at risk at lower tier assessment steps. Here, simulation models serve as an important function for risk assessments because they provide the mechanism to conduct this translation consistently and transparently, allowing for emergence of the case-specific relationships between the fates of individuals and the populations under consideration (Schmitt et al. 2015).

What protection goal is being addressed?

Various protection goals can be addressed, e. g. no decline in population size or negligible effects on population structure.

What outputs are required?

Population development over time to assess trends in population size. Occupancy would be a good indicator of separating the overall population size effect into spatial units, i.e. some habitats might be more affected than others, and source-sink structure within a population could be identified. Population age structure, i.e. the relation of adults vs. young is also a useful indicator of population health.

How was the species chosen?

ALMaSS has been applied to different small mammals and terrestrial invertebrates, with specific adaptations in the model formulas to meet the life cycle of each species. In this review, two applications for the vole and the brown hare are considered (Schmitt et al. 2015, Topping et al. 2016). For the simulation of the vole and wood mouse in both ALMaSS and (Liu et al. 2014), a well investigated small mammal was selected that inhabits agricultural areas and is -through herbivory - directly affected by land use and application of PPPs (so called indicator species).

Which other species/groups are being covered by the chosen one(s)?

None, because the parameters and processes of the single species models are very specific. However, the ALMaSS framework can be used for any species.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Long-term population data of a respective species are sometimes used, depending on the modelling application. For the field vole application of ALMaSS, a visual interface was used and vole ecologists verified plausibility of results as well as the pattern-oriented modelling (Topping et al. 2012) for age and sex structure, vole densities across multiple habitat types, dispersal, and emergence of realistic predator-prey cycles. For the brown hare and skylark applications of ALMaSS (Topping et al. 2010b, Topping et al. 2013), a 13 years' time series of population dynamics was used. So far, there is no 'gold standard' assessing whether e.g. a real time-series matches or not.

Regarding the wood mouse, Liu et al. (2013) conducted pattern-oriented validation (Railsback and Grimm 2011) on three separate patterns: population dynamics, habitat usage and proportion breeding. The population dynamics with default parameter values and no pesticide exposure were compared with data from field experiments. They also compared the proportion of the populations that was breeding in the model with observations from field studies of Elton et al. (1931) and (H. Tattersall et al. 2004). Following criteria for comparing model outputs with field data were used: general patterns of the population time series from peak to trough, overall levels of population development and seasonal differences rather than precise matches of population numbers.

2.9.2.2 Supporting Data

Summary of the key data used in the model for development and evaluation

The parameter values for the species are based on literature: sometimes, best guesses have to be used (e.g. habitat suitability scores for different land use types). The parameters for the PPPs and their effects are related to known effects in other species under lab conditions, but remain so far fictive both in terms of the effects on the actual species as well as effects under field conditions.

Assessment of the quality of the data

As data from peer reviewed literature was used for the species' biology, the quality can be judged to be fairly good. How far case studies from certain areas are applicable to the respective setting of the ERA application needs to be assessed separately.

2.9.2.3 Conceptual Model

Description of the model concepts including a diagram

See general text above for description and diagrams: the landscape, treatment and individual-based sub-models are connected in such a way that energy uptake defines foraging and dispersal in the IBM. That is, the animal has to fulfil its energy requirements and therefore searches for suitable patches to forage until these requirements are met. Also, movement steps are linked to energy expenditure and define dispersal distances. Ultimately, also reproductive success is linked to the animal's energy budget. These processes, together with the pesticide treatment and the configuration of the landscape (defining suitable habitat), this estimates the uptake of pesticides. In recent applications, the effects of a fictive PPP are translated into (plausible, but still not validated) effects on individual's reproduction, which feeds back into the population level output of population development and densities over time in a given spatial context. The framework is flexible, so that any other output can be retrieved, e.g. population structure, population spread etc.

Identify the main components and processes in the system

See above: The main components are a landscape model including pesticide treatment and an animal IBM with their respective processes. Depending on the application, any situation and scenario can be included. E.g. realistic, spatially-explicit landscapes and farming practices (e.g. for the wood mouse with annual crop rotation) can be included at a 1 m^2 scale and spatially-explicit treatments as well as their repeated pulses can be included. The IBMs are also flexible and include the basic processes foraging and growth, dispersal, reproduction and survival, with these basic processes interfering with the landscape (e.g. foraging, dispersal) and the treatment sub-models (survival, reproduction).

How the effects of the chemicals are modelled

A standard Tier 1 risk assessment is performed in the following way (citation from Schmitt et al. 2015): "A synthetic toxicological dose–response relationship reflecting realistic worst impacts on reproductive performance is constructed based on the effect profile derived by collecting toxicological results from 15 different pesticides (Mastitsky et al. 2013)".

In ALMaSS, both lethal and sublethal direct effects of pesticides are considered, whereas Liu et al. (2014) considers only mortality. In both models, the dose-response relationship is calculated based on the application rates, derived residues in food, and the species-specific daily food intake. In a case study using both models (Schmitt et al. 2015), the residues were assumed to decline over time with a

standard half-life of 10 d, and the daily dietary dose (DDD) was calculated using the time-weighted average concentration over 21 d (DDD21d). From this, toxicity-exposure ratios were calculated.

In the wood mouse model (Liu et al. 2014), the only direct effect of pesticides considered is mortality caused by the ingestion of treated seeds. The mortality is either based on the internal pesticide concentration within the body, as calculated from a toxicokinetic (TK) sub-model using absorption (k_a) and elimination (k_e) rate constants. Alternatively, the mortality can be based on the body weight-normalised daily-eaten dose of the pesticide, excluding TK modelling. Following the widely used algorithm "individual tolerance" (Jager et al. 2011), individuals die when their internal concentration (or the daily consumed dose) exceeds an innate tolerance concentration. The authors assumed that parent mice do not pass the toxicant to infants during gestation or lactation. As a result, infant mice are not exposed to the pesticide in the model, but they die without parental care if the mother has died.

When TK is included, the stepwise changes (i.e. discrete time step: every 1 min) in gut (C_{gut}) and internal ($C_{int}t$) concentrations of pesticide are calculated under the assumption that all toxicant is bioavailable (i.e. 100% absorption from the gut into the system) as follows (Bednarska et al. 2013):

$$\Delta C_{gut} = I - k_a * C_{gut}$$

$$\Delta C_{int} = k_a * C_{gut} - k_e * C_{int}$$

where ΔC = change in the concentration of pesticide (i.e., body weight-normalised dose of pesticide) in given time interval, here one minute; I = ingestion rate (i.e. the rate of toxicant transfer from exposure dose to the gut, mg a.i./kg body weight/min); k_a = absorption rate (per min) and k_e = elimination rate (per min).

For comparison, when absorption and elimination are not considered explicitly, the exposure is represented as daily eaten dose (D_{eaten} =sum(I)). Depending on the food intake and pesticide application rate, pesticide accrues in the animals within a day. As the authors assumed that the hypothetical toxicants were fully excreted before the next nightly feeding, in both TK and non-TK calculations the exposure levels are re-set to 0 in the beginning of each day.

By incorporating absorption and elimination rates and the exposure duration (1 min was assumed for gavage exposure for comparison with acute toxicity study), the dose–response curve for D_{eaten} can be translated to the concentration–response curve for C_{int} (Table 12). Based on this concentration-response curve, the acute endpoints (i.e. individual tolerance dose and concentration, above which the animal dies) are defined specifically for each individual mouse at birth, i.e. x% of mice have tolerances that are lower than LDx (or LCx, x is in the range of 0–100). The tolerance is calculated once and remains the same throughout the lifetime of a wood mouse. The response curves are implemented via a probit function (Table 12). In the wood mouse model of (Liu et al. 2013), Liu et al. (2014), only D_{eaten} or C_{int} exceeding lethal thresholds on a daily basis are considered for acute effect (i.e. mortality). However, there is no reference to literature in the mentioned Table 12 for the parameters of the dose-response relationship.

In the brown hare model using ALMaSS (Topping et al. 2016), a TD-like sub-model is implemented to assess the effect of an endocrine disruptor on the respective mammal. This sub-model is based on a kind of energy budget, i. e. a first-principle approximation of how energy (food resources) is triggering other processes like movement and search for food, as well as reproductive success and survival. The model includes internal and external toxicokinetics in terms of the varying rates of ingestion of the pesticide, and the process of elimination within the hare. The internal TK are represented by a single compartment model assuming a percentage elimination rate per day. External TK is determined by the feeding behaviour of the hare and ultimately by the time spent feeding from contaminated areas, and by the concentration of the pesticide in the vegetation. This resembles a very basic effect model: "For the chronic effect, the impact of exposure above a threshold body-burden is modelled as a uniformly distributed chance of litter size reduction of 0–100% for female hares exposed during gestation of that

litter" (Topping et al. 2016). In other words, there is only a threshold, and above that threshold, effects are no longer dose-related but random. The effect used was intended to represent a chronic reproductive effect over time.

In another application of ALMaSS for voles (Dalkvist et al. 2013), the effect of the fungicide vinclozolin is modelled. The pesticide exposure was modelled as a single foliar application to the grass between the trees in orchards once a year. The pesticide residue concentration on the vegetation was recalculated every 24 h based on the pesticide's half-life of 7 days, until its concentration fell below 0.00001 mg/m^2, thereafter it was assumed to be zero. Drift to off-crop areas was modelled based on drift data from the spray drift calculator within FOCUS surface water scenarios (FOCUS 2001). The amount of pesticide consumed by the voles in and around the treated areas was affected by the level present on the vegetation, the weight of the vole and the rate of ingestion, and was calculated by the use of a standard index used in the field of risk assessment as:

$$Pesticide\ intake = I * P * V$$

where I is the ingestion rate (1.39 kg food/kg body weight), P is the pesticide concentration (mg/kg) and V is the typical weight of the mother (25 g) (Crocker et al. 2002). Bioaccumulation of vinclozolin has not been documented (http://www.inchem.org) and was consequently not implemented in the model.

How the components and processes are linked

The basic IBM processes foraging and growth, dispersal, reproduction and survival interfere with the landscape (e.g. via foraging, dispersal) and the treatment sub-models via survival and (in ALMaSS) reproduction.

2.9.2.4 Formal Model

Identification of the model variables

This depends on the application and the species modelled. See general introduction for an overview over model parameters and variables for the wood mouse example.

Identification of the model parameters

This depends on the application and the species modelled. See general introduction for an overview over model parameters and variables for the wood mouse example.

Description of the most important model equations or algorithms

They are specific for each application. See section 2.9.2.7 for modelling the dose-response curves, TK and concentration-response curves.

2.9.2.5 Computer Model

Description of the model implementation

The model is implemented in C++. The documentation is well structured, follows the ODD guidelines with own ODdox pseudocode standard and a website with demo. There is an interface with the free software R to import and create landscape layers (package 'ralmass' on github: https://github.com/LDalby/ralmass).

Checking the computer model for errors, bugs and inconsistencies in the code

Each implementation has its own development history, but e.g. the field vole model has a long history of development (Topping et al. 2003), application (Jepsen et al. 2005, Dalkvist et al. 2009) as well as testing and calibration (Topping et al. 2012). In most ALMaSS publications, robustness/ sensitivity analyses have been conducted and a plausibility check of the results has been made ('informal' in most publications, see section 2 supporting data). The code is open source, but the information to which extent it has been crosschecked independently is not available.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

See above. A formal testing and fine-tuning to field data has been carried out in the publication of Topping et al. (2012) for the field vole.

2.9.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context **in which the model is run**

Flexible and context-dependent. The size, composition and structure of the landscape as well as its temporal change, e.g. due to agricultural management or seasonal changes, can be adapted. Scenarios can be chosen in such a way that they are conservative in respect to the effects caused by the pesticides to be assessed (e.g. field voles: spray application in orchards; wood mice: seed treatment in cereals; see examples below).

Field vole example (Schmitt et al. 2015): Three realistic landscapes containing 2.5% of treated orchards, where scenarios are named for the actual locations. The non-orchard crop areas are simulated as being managed following normal farming practices, but without applying any pesticides toxic to voles. However, because field voles depend on well-developed vegetation cover, there was very little alternative habitat (other than the orchards) in the landscape, so that basically the vast majority of all field voles in the landscape inhabited the treated orchards.

Wood mouse example (Schmitt et al. 2015): Four fields with rotating crops (10 ha each) surrounded by 5 m wide hedgerows. This actually represents a rather good situation, i.e. not a vast monoculture landscape. Crops rotate yearly. In any case several years of consecutive use of the substance plus a reasonable postexposure period should be simulated. Usually 10 years with applications should be a reasonable period.

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

Parameters of spatial scales (1 m^2) and temporal scales (1 day) simulated over several generations of the focal species fit with the important processes, e.g. foraging and pesticides uptake, effects on survival and reproduction, population spread and dispersal to suitable habitat as defined by the landscape structure.

2.9.2.7 Parameter Estimation

Description of the model parameter estimation

In the published model demonstration for the model of Liu et al. (2014) and in the wood mouse application of ALMaSS (Schmitt et al. 2015), the fungicide is fictive, and the effect on reproduction is based on a plausible relationship of body weight at reproduction. A statistical evaluation of data from literature used to derive the toxic properties of the fictive fungicide showed a certain correlation between

body weight of mothers and number of offspring. However, no clear and quantified relationships between body weight (bw) and relevant life history processes are reported in the literature. A synthetic toxicological dose–response relationship reflecting realistic worst impacts on reproductive performance was constructed based on the effect profile derived by collecting toxicological results from 15 different pesticides (based on a publication by Mastitsky et al. (2013)). For each endpoint, the 33%, 66%, and 95% quantile of the magnitude of the effects observed over all substances was determined and assigned to the lowest, middle, and highest dose, thus yielding a generic realistic worst-case dose–response relation.

Table 13: ALMaSS / Wood Mouse Model – Toxicological Profile in Case Study Risk Assessment

	Parental F0 females			
	Dose level (mg/kg bw/d)	Body weight at mating	Body weight at birth of F1	Body weight at weaning of pups (day 28 pp)
		% control		
NOAEL (low dose)	20	100.4	100.1	100.7
LOAEL (mid dose)	50	97.6	96.9	97.7
High dose	150	91.7	90.7	89.7

	F1 pups					
	F1 Number at birth (day 0)	F1 Number at weaning (day 28 pp)	F1 Body weight at birth (day 0)	Body weight at weaning (day 28 pp)	Delay sexual maturation females	Delay sexual maturation males
	% control				days	
NOAEL (low dose)	101.2	100.0	100.0	98.3	0.3	-0.3
LOAEL (mid dose)	92.9	96.2	96.6	93.2	1.0	0.7
High dose	71.8	71.1	86.9	67.8	2.4	2.2

Data from Schmitt et al. (2015, edited).

Parameters estimated from the literature — what are the sources and why are these appropriate?

A simplified approach was used to derive dose–response functions. Effects were set to zero for doses below the NOAEL (no observable adverse effect level; see table above with toxicological profiles based on a summary of Mastitsky et al. (2013)), and above that level a linear function was fitted to the experimental data (see Table herein), with up to 100% effect in the high dose range. This dose–response relation was then used in combination with the simulated exposure to derive the model parameters for individuals in the affected populations (Schmitt et al. 2015).

Table 14: ALMaSS / Wood Mouse Model – Fitting of Individual-Level Effects

	Effect = *a* * Effective dose + *b*		
Coefficient	Litter size (% control)	Pup survive (% control)	Delay of 1st reproduction (days)
a	-0.0022	-0.0023	0.0156
b	1.0494	1.0593	0.0885

Coefficients a and b for the linear dose-resonse curves of the three types of sublethal effects when 20 mg a.i./ kg bw < effective-dose <= 150 mg a.i./kg bw. Data from Schmitt et al. (2015, edited).

Parameters obtained from calibration — how and why this was done?

The model demonstration in Schmitt et al. (2015) shows a pure application, no calibration was performed. Since so far, no data exist on dose-response relationships for the respective species, a calibration of the regulatory model cannot be conducted.

2.9.2.8 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

Sensitivity analysis/ robustness analysis and uncertainty analysis are used synonymously here. As described above, scenarios with low, middle and high doses were tested. Results cannot be generalized, as they depend on the respective species, the amount of untreated habitat in the landscape, patch size (in large fields, animals do not reach the refuge habitat etc.). The variation linked to both farming and landscape makes matching the correct farming with the correct landscape critical to prediction of impact (Topping et al. 2016).

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

See above.

2.9.2.9 Comparison with Measurements

Description of comparisons of model output with independent data

For the regulatory model, no measurements are available.

Demonstration that the model output provides an adequate match to data patterns

For the regulatory model, no measurements/ pattern are available (but see later in section 2.9.3.5 the pattern match for the complete model performance).

2.9.2.10 **Model Use**

Explanation of how the model conforms to the requirements set in the problem definition

The model is well suited to derive predictions at the landscape and population scale, with the quality of predictions certainly depending on how good the model performed under plausibility tests and 'evaludations' with other data or independent data.

Description how the model works (user manual).

Regarding ERA, demonstrative applications of the model have been published for skylarks (Topping and Odderskaer 2004, Topping et al. 2005), beetles and spiders (Topping et al. 2014, Topping et al. 2015), voles (Dalkvist et al. 2009) and brown hare (Topping et al. 2016). ALMaSS comes along with a very good pseudo-code documentation of processes and parameters following the ODD (Grimm et al. 2006):

http://projects.au.dk/casesm/almass/

http://www2.dmu.dk/ALMaSS/ODDox/ALMaSS_ODDox/V1_01/main.html

Description of the pesticide parameters values used in the model

The pesticide parameter values used in the model demonstration of ALMaSS and the wood mouse model of Liu et al. (2014) by Schmitt et al. (2015) are reported in the tables of Schmitt et al. (2015).

Description of the specific assessment including a discussion of the most important results

Context dependent and provided in the specific results and discussion sections of Schmitt et al. (2015).

2.9.3 Model Evaluation

2.9.3.1 **Problem Definition**

The regulatory context in which the model is run

For vertebrates, "no individual mortality" has been suggested as short-term protection goal (EFSA PPR 2014b). In contrast, the focus of the long-term risk assessment is on reproductive effects as measured in toxicological studies in rats or rabbits. ALMaSS and Liu et al. (2014) can be run in this context to study whether reproductive effects translate into long-term population declines.

The question that has to be answered with the model

Questions that can be addressed with both models in Higher Tier risk assessment include the prediction of long-term effects of a pesticide on the population size of a terrestrial vertebrate indicator species in a spatial context and under various external stressors (harvesting, climatic effects). Since the fate of each individual can be followed, other measures of population health, like age structure, can also be yielded by the model.

The available knowledge and data relevant to the risk assessment question

The modelled effects of pesticides on the individuals are based on literature on laboratory model species (e.g. rats, rabbits), i.e. not the focal animal (e.g. vole or brown hare), because such measurements are lacking. Physiological reactions in rats and rabbits might strongly differ in relation to voles and brown hares. It is clearly stated by the authors that the toxicological effects assumed were entirely constructed; fertility reduction is the ONLY effect considered! There might be other effects e. g. on survival and longevity, behavioural consequences like disorientation leading to increased mortality. The results are not intended to represent any real pesticide, but were designed to demonstrate noticeable impacts of a chronic reproductive effect over time (Topping et al. 2016). Please note that the TK part has NOT been validated for each respective species and is based on fictive (yet plausible) effects of a FICTIVE pesticide.

The outputs required to answer these questions including performance criteria for the regulatory model

The required output (population performance over time) is clearly addressed.

The species to be modelled

The models are intended to cover small mammals by addressing the field vole and wood mouse as model species. A general question is the usefulness of a geographically widespread species, that is thriving in cultural landscapes and does not seem to be affected by current land management practices, as a representative species for ERA to represent vulnerable species of the 'small herbivorous mammal' guild. There seems no literature stating population declines in field voles since the 1980ies where agricultural practices changed and were intensified. The brown hare or skylark seem to be more appropriate species for this purpose.

Requirements for the environmental scenarios to be used in the risk assessment

Requirements are clearly stated (i.e. farming and treatment practices).

2.9.3.2 Supporting Data

Are the data fit for purpose in view of the problem definition?

Each model component comes with a detailed description of parameters and processes that are published in peer-reviewed journals; the spraying/ PPP application regimes in space and time are clearly defined.

Has the quality of the data used been considered and documented?

For the field voles, authors are referring to Danish agricultural landscapes. Authors have, where possible, used parameters that have been measured from non-cyclic Scandinavian populations to closely simulate the Danish vole population (Topping et al. 2003). They have used independent data to validate four different population patterns with the model (Topping et al. 2012): age structure and density, vole density in multiple habitats, vole dispersal, and vole population cycling. However, from the publications it seems that the independent data for validation did not stem from populations in non-treated (pesticide-free) agricultural fields. Thus, the validation presumably comprises the effect of pesticides, therefore the quality of the data may be judged high for this example.

Have all available data been used? If not, is there a justification why this information has not been used?

It cannot be assessed in this review whether all available data has been used; this would need, for each modelled species, detailed expert knowledge. *M. agrestis* is one of the most well studied small mammals with hundreds of studies covering molecular ecology, behavioural ecology, predation, feeding ecology, habitat selection, and cyclic dynamics in particular (Topping et al. 2012).

2.9.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

Yes, these are population numbers/ densities over time.

Are the modelling endpoints relevant to the specific protection goal?

Yes, individual-level endpoints derived from toxicological studies, e.g., body weight of parents and offspring, litter size and viability of pups were translated into population level effects to show by way of example a compliance with the protection goals.

Is the modelling approach justified?

The assessment of population dynamics prerequisites the knowledge on the drivers of density regulation, i.e. from bottom-up (habitat / resource driven) to top-down control (predation, competition). As ALMaSS is based on space and thus includes habitat and forage availability, the underlying biological model (field vole model) bears elements of first principle rules (energy reserves and food uptake), basing behaviour (movement, dispersal) and demographic traits (reproduction, mortality) on resource availability (bottom-up control). Hereby, as in any individual-based models, assumptions have to be made, e.g. on classifying habitat quality via scores (Table 5 in Topping et al. (2003)):

Table 15: ALMaSS – Vole Habitat Quality Categories

Type	Vegetation characteristics	Typical landscape elements	Habitat score
Optimal habitat	Cover > 80 % and height > 40 cm	Non-grazed uncut grassland, old set-aside, field and road margins	3
Sub-optimal habitat	Cover > 40 % and height > 10 cm	Young tree plantations, dry meadow or heathland, cereal crops, undersown with grass	2
Marginal habitat	Cover < 40 % or height < 10 cm	Grazed or cut grassland, cereal and grass crops	1
No habitat	No grass	Roads, mature tree stands, other areas with no grass	0
No habitat	None	Buildings, water bodies	-1

Parameter values were adapted from Hansson (1977) for Danish conditions and habitat types available in ALMaSS. Habitat scores are an arbitrary classification. Data from Topping et al. (2003, edited).

Bottom-up density regulation, apart from dispersal, is reached when spending too long time in unsuitable habitats (accumulate too many starvation days), or at the end of the individual's life-span. If the dead vole is female it is assumed that any non-weaned young would also die. Top-down regulation happens via being killed by a predator (here implemented as a background mortality and arbitrarily set to 2%), or other external factors (e.g. farm operations). "Agricultural mortalities" were given three values depending upon the type of agricultural management. Pig grazing mortality was 25% per day, harvest and cutting mortalities were 50% and all soil cultivation was deemed to result in 75% mortalities. In all cases, only those voles present on the field under management at the time the management was carried out are subjected to agricultural mortalities (Topping et al. 2003).

While the bottom-up part seems quite well thought of, top-down regulation is not (yet) well developed. As ALMaSS so far provides single species models, there is no competition (niche overlap) with other animals included, which might get decisive when the dominant species gets controlled by a pesticide agent and subdominant species are less sensitive to pesticide application. This could seriously increase time to recovery of the focal species' population or even lead to population decline or extinction. Also, predator-prey relationships are only modelled via an arbitrary percentage; there is no adjustment procedure e.g. under predator release.

Is the conceptual model logical?

Yes, but depending on the species and feedbacks, it can be quite complex.

Are the processes included in the model relevant to the addressed issue?

All in all, it is difficult to assess for a non-expert whether all important processes have been included; e.g. predation pressure might only play a minor role in the vole life-cycle in Danish landscapes. As another example, harsh winters with associated high mortalities are not considered.

Are the links between different processes to the variables logical?

Yes, see above model description above.

Are the temporal and spatial scales relevant in regard to the problem definition?

Yes, see above model description above.

2.9.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

Description of the formal models is given in the ODD-protocol (Grimm et al. 2006), and the models are accompanied by pseudo-code documentation (see for example http://www2.dmu.dk/ALMaSS/ODdox/Field_Vole/V1_02/index.html), flow charts with process scheduling as well as detailed parameter description and values along with their justification in published results.

Are the most important mathematical equations described?

Yes, see above model description above.

Is there a description of the variables and parameters including their meaning and unit?

Yes, see above model description above.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

Since the field vole is a common and wide-spread species, the formal biological model seems well justified with peer-reviewed literature (but see shortcomings in modelling population regulation above, 2.9.3.3) and has been used in a variety of tests and applications (Topping et al. 2003, Dalkvist et al. 2009, Topping et al. 2012, Dalkvist et al. 2013).

The European brown hare model was built based on available literature data, using multiple field data patterns from an island hare population study to guide model development. Application of the model to a comprehensive historical data set supported a functional relationship (Topping et al. 2010a).

Are references supporting the equations been provided?

Yes, see model description above.

2.9.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

ALMaSS comes along with a very good and transparent online-available code-documentation (ODdox documentation created by combining a modified version of the ODD protocol with documented code using pseudo-code doxygen (van Heesch 1997); http://www2.dmu.dk/ALMaSS/ODDox/Field_Vole/V1_01/index.html).

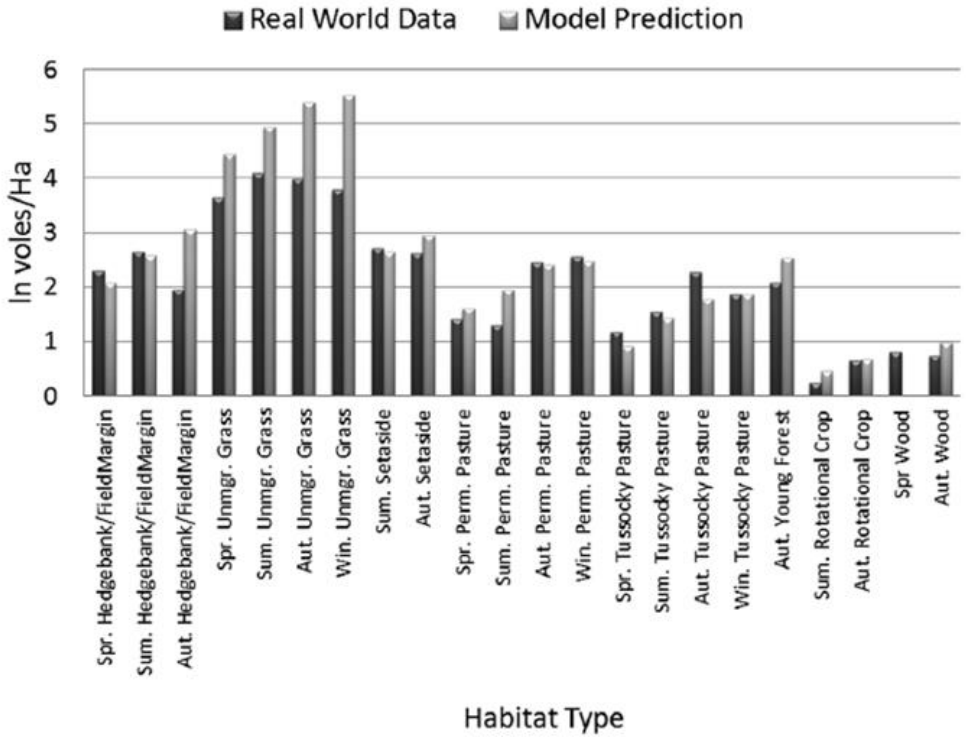
Is the computer code well readable and is it available?

ALMaSS has been programmed in C++ and has a user interface; however, for access to the source code and the ALMaSS platform contact to the owner and its institution is recommended, although ALMaSS is an open source project hosted on the Collaborative Computing Projects site CCPForge (http://ccpforge.cse.rl.ac.uk/gf/project/almass/; now moved to GitHub), where code and further documentation are hosted (Dalkvist et al. 2013).

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

Since the field vole / brown hare models have been 'evaludated' with independent data, it can be concluded that a reality check has been applied for this species' biological sub-model (see Fig. 18 below) (but see shortcomings in modelling population regulation above). The predictions are quite matching reality, however, in unmanaged grasslands the deviations can be up to 30%, especially in autumn and winter, indicating that seasonal mortality might not be accounted for correctly.

Figure 18: ALMaSS – Validation of the Vole Model



Real world means and model means for total vole density for a range of Danish habitats and sampling periods. X-axis abbreviations: spring (Spr.), summer (Sum.), autumn (Aut.), winter (Win.). Unmanaged (Umgr.), Permanent (Perm.) Graph reproduced from Topping et al. (2012).

2.9.3.6 Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

ALMaSS is a platform, any landscape at any scale can be plugged in; for the field vole, parts of the Danish agricultural landscape have been used that seem highly structured and diverse.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

A better scenario would be using a reasonable worst-case, i.e. an agrarian steppe without many hedge rows and low habitat quality. The 'realistic' landscapes might still contain too many refuge areas.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

Whether the exposure scenario is realistic is difficult to evaluate when the pesticide is fictive, however, the deduction of the exposure scenario from literature seems to follow what is known from effects of

other pesticides on organisms. See model description above for a detailed list of how pesticide effects were modelled.

Is the level of conservatism placed into the scenarios appropriate?

Yes. Scenarios are modelled over decades and the landscape is sufficiently large.

2.9.3.7 Parameter Estimation

The model parameter estimation has been adequately documented?

Please mind that the experiments to assess the parameters of the regulatory model are lacking at this stage. The authors have done their best in assembling knowledge from other Tier 1 studies on rats and rabbits and transferring this knowledge into a TK-TD framework (e.g. brown hare model), but Tier 1 studies on the focal species are missing. Further, it is unknown whether reproduction reduction is the only effect in the host species, or whether other effects are additive. Please refer to model description on how the Tier 1 studies have been applied.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

So far, no literature or experimental data exist to calibrate the effect model.

Were the estimated parameter values realistic?

This cannot be evaluated - see points above.

Are the data sources sufficiently documented?

Yes.

2.9.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

A sensitivity analysis in ALMaSS has been carried out on the underlying biological vole model and consisted of visual inspection of the output at higher levels (population dynamics) as well as changing parameter values. It is adequately documented, see below for an example of Topping et al. (2003, p. 78).

The model analysis was carried out in two phases. The first phase (plausibility) consisted of using the visual interface of the model to monitor individual vole behaviour in the landscape and to use vole ecologists to verify this. The second phase was a sensitivity analysis and comprised altering the main vole parameters by ±5, ±10 and ±20% and assessing the effect on the population. This procedure identified habitat quality (HQT 1–3 from low to high) to be the most sensitive parameters. In particular, changes in HQT 1 were found to result in correspondingly large changes in population size. This is not surprising since these parameters describe the fitness of the vole. Other parameters such as increased reproductive success only increased population size when carrying capacity as set by HQT was not reached.

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

Yes, depending on the context and the application, the effects of the pesticide or the spatial treatment scenario are varied.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

Yes, they are mostly presented in result tables.

Has the uncertainty analysis been adequately documented?

There is no difference made between sensitivity/ uncertainty/ robustness analysis, hence the above mentioned is valid.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

See above.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

See above.

Uncertainty is propagated to the model results?

See above.

Have confidence intervals been estimated and has this information been used in further model use?

This is dependent on the application. E.g. in Schmitt et al. (2015), lower and upper limits for dose-response-effects have been derived and used.

2.9.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

Please refer to point (5). Yes, the vole model has been compared with independent data and further calibrated. The performance in habitats like unmanaged grasslands showed deviations up to 30%, especially in autumn and winter, indicating that seasonal mortality might not be accounted for correctly in these habitats or due to overestimating habitat quality.

Are the model outputs that are compared relevant in view of the problem definition?

Yes, population time series are the measured endpoint.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

Yes, these are long-term field data of population censuses. However, secondary literature on how population censuses were obtained have not been reviewed here.

Is the dataset relevant in view of the problem definition?

Yes, as population development over time is the endpoint, real time series are relevant.

Is the fit of model output to the data good enough?

This is difficult to assess, as there are no criteria defined how well they should fit. Despite detailed pattern-oriented model fit (Topping et al. 2012), deviations in time series up to 30% were reported.

Has the performance of the model been reported in an objective and reproducible way?

Yes, time series plots of model output and real data are available.

2.9.3.10 Model Use

Is a user manual available?

Yes, ALMaSS is designed for transparency in modelling and model testing, to facilitate the reproducibility of scientific results, for freely available source and public availability and reusability of scientific data, and for public accessibility and transparency of scientific communication (https://ccpforge.cse.rl.ac.uk/gf/project/almass/). All code is available also in GitLab containing all the code for both the command line and GUI versions of ALMaSS (https://gitlab.com/ChrisTopping/ALMaSS_all); see also Introduction to ALMaSS following good scientific modelling practice/ ODD and application to ERA. For using ALMaSS, a request needs to be sent to the authors to join the project.

Have all aspects of the modelling cycle been documented?

Yes, an ODD standard is inherent to ALMaSS and the Wood Mouse Model of Liu et al., which is basically containing the elements required under section 11.2 of the EFSA Sci. Op. on GMP (2014b).

Has a summary sheet been provided by the modeller?

A standardized summary document was not found.

2.9.3.11 Suitability of the Model for Regulatory Purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

Yes, a dialogue between modeler and risk assessor is possible.

Is a version control system implemented?

As the code is on GitHub, this automatically contains a version control.

2.9.3.12 Overall Judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

Overall, ALMaSS may be judged to be on a good way to become an appropriate, maybe even prime modelling platform for risk assessment, as many processes of the biological model are based on first principles (animal physiology and energy handling leading to emerging foraging behaviour) and the spatial application in a landscape and habitat context might be very helpful for ERA and opens new doors of thinking, as these aspects are usually not included in ERA.

Further, ALMaSS and the Wood Mouse Model are important upscaling tools from laboratory conditions to effects of repeated annual pesticide applications in a spatial context. The models can be considered as most important steps forward towards increasing realism and the capacity of integrating non-linear feedbacks and mechanisms. For future use of ALMaSS, if an acceptable risk is demonstrated at lower tiers, it would be useful to crosscheck that the pesticide is still safe at the landscape level, i.e. ALMASS should be used as an additional check that protection goals are met (e.g. as in the EFSA Sci. Op. on non-target arthropods, EFSA PPR 2015a).

However, more effort is required regarding the implementation and validation of the toxicity module for direct (individual-level) effects. TKTD models in Higher Tier studies should meet quality criteria that are at least similar to those that have been defined for the lower Tier 2C in the aquatic risk assessment. Therefore, if a TK part is used, then the recommendations of the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) should be followed regarding validation, i.e. at least two different exposure profiles with three concentrations triggering effects in the range of partial effects should be used.

With the drawbacks mentioned, so far, the strength of ALMaSS lies in its scientific value of understanding complex dynamics and interactions of individual behaviour, habitat configuration and spatial food availability, together with farming practices and spatial pesticide application.

The drawbacks so far are:

- No Tier 1 risk assessment available for model set-up, thus the validation of the TK(-TD) part for each model could therefore not be done yet.
- The toxicological effects on reproduction are not validated. Other aspects on physiology (body size, growth rate) are not addressed, maybe there are more sensitive endpoints like survival (expressed via body weight in the F1 generation, but they would not survive a harsh winter = additive effects). What happens under multiple exposure cannot be addressed.
- For most models, a rigorous testing and validation is missing, simply also because long-term field studies with measurements on individuals are missing. Thus, there might be crucial, but unknown processes (as is the problem with modelling in general)
- Field voles/ wood mice might not be the appropriate species for ERA as they are widespread and hemerophilous (commensals of cultivation). The brown hare example is more appropriate, but so far less investigated (only 1 published application example).
- A deviation of 30% (vole model) or even 100% (wood mouse model) from measurement endpoints, here population density, in the validation is maybe an indication that important processes are missing, or that there is so much innate variance in the natural processes that it will be extremely difficult to tease apart the effects of pesticide application – at least for robust species.
- A weak part is the missing community context of the model (predation risk, competition and niche overlap) that has not been considered at all. This is not a flaw of the developers, but of a general lack of individual-based models in a community context, that only recently get developed. I also see the difficulty in ever being able to parameterise these models at all, as it would mean quantifying niche overlap between all species in a guild as well as functional responses in predator-prey dynamics, such as effects of predator release and prey switching. All these processes will influence population growth and density regulation of the focal species. But the development and testing of these models will take ages, and adding more complexity will also increase the risk of error propagation and not necessarily lead to higher precision.

Therefore, risk assessment should be based not only on one model output, but on an ensemble modelling approach as in other fields of science (e.g., species distribution modelling in conservation ecology) where possible. That means, the results of several modelling approaches should be compared; where results are in agreement, confidence in the model results could be higher, and where results deviate, this should lead to exploring why. That is, deviations might hint to crucial processes and mechanisms

not included in one modelling approach, but that are essential to describe the system's dynamics. This will further the development of predictive models in ERA.

To further IBMs in ERA, randomised before–after control-impact (BACI) studies should be initiated and results compared to the model output, together with furthering Tier 1 studies, to refine and tune the models that are currently available.

2.9.4 Qualitative Assessment of Uncertainties

2.9.4.1 **Potential for Underestimation of Real Risk**

- No Tier 1 risk assessment available for species model set-up, thus a formal validation of the TK(-TD) part for each model could therefore not be done yet. However, this is not a flaw of the models but of a general lack of long-term field data and experimental setups.
- The toxicological effects on reproduction of the modelled species are not validated. Other aspects on physiology (body size, growth rate) are not even addressed, maybe there are more sensitive endpoints influencing survival, e.g. expressed via body weight of the F1 generation. Possibly F1 offspring might then not survive a harsh winter, displaying additive effects. Also it is not clear what would happen under multiple exposure
- Field voles/ wood mice might not be the appropriate vulnerable species for ERA as they are widespread and hemerophilous (commensals of cultivation). The brown hare example is more appropriate, but so far less investigated (only 1 published application example).

2.9.4.2 **Potential for Overestimation of Real Risk**

- See below.

2.9.4.3 **Potential for Uncertainty in Either Direction**

- A weak part is the missing community context of the model (predation risk, competition and niche overlap) that has not been considered at all. This is not a flaw of the developers, but of a general lack of individual-based models in a community context, that only recently get developed. It is questionable if these models can be appropriately parameterised given all the uncertainties under field conditions, as it would mean quantifying niche overlap between all species in a guild as well as functional responses in predator-prey dynamics, such as effects of predator release and prey switching. All these processes will influence population growth and density regulation of the focal species. But the development and testing of these models will take a long time, and adding more complexity will also increase the risk of error propagation and not necessarily lead to higher precision.
- The missing validation with independent field data leaves it unclear, whether all important biological processes are included for the different species.
- Environmental variability.

2.10 $SPEAR_{pesticides}$

Evaluation by Matthias Liess

2.10.1 General Information

2.10.1.1 Background and Concept

SPEAR (Species at Risk) is a spatially explicit, trait-based family of models developed to link the toxic pressure to the ecological effects of toxicants in streams. Endpoint is the proportion of vulnerable invertebrate species. SPEAR can be applied retrospectively to identify the toxic pressure at a site when the composition of invertebrate community is known; SPEAR can also be applied prospectively to predict the composition of invertebrate community when the toxic pressure at a site is known. Various SPEAR approaches were developed to identify the long-term effects of different types of toxicants and their respective patterns of exposure:

$SPEAR_{pesticides}$: Links agricultural pesticide exposure to invertebrate community structure. The approach was developed by Liess and von der Ohe (2005)[20]; the latest version has been described in Liess et al. (2021)[21]. The following document will focus on this member of the SPEAR family.

$SPEAR_{metals}$: Links metal exposure to invertebrate community structure. The approach was developed by Liess et al. (2017).

$SPEAR_{oil}$: Links bitumen-derived contaminants from oil sands to invertebrate community structure. The approach was developed by Gerner et al. (2017).

Realistic Consideration of Contamination

Agricultural pesticide exposure happens through surface water runoff and related rainfall induced fast processes (Liess et al. 1999), slower pesticide transport through drainages and direct overspray (Ganzelmeier et al. 1995). Generally, the fast processes are most relevant for the less soluble substances with high insecticidal activity (Reichenberger et al. 2007). For the design of $SPEAR_{pesticides}$, the resulting short-term peak concentrations were obtained using Event Driven Samplers (EDS) (Liess and Schulz 1999, Liess and von der Ohe 2005). In brief, EDS consisted of two 1 L glass bottles fixed on a stainless-steel rod with an opening at 5 cm of the lower bottle and approximately 10–15 cm of the second bottle above the water level, depending on the normal expected rise in water level, which was determined empirically at each sampling site. The bottles filled with water as the water level rose after rainfall events of at least 10 mm/day. Bottles from EDS were checked within 48 h after each rainfall. With this approach the real contamination occurring in the environment is estimated in the most realistic way. We identified that the highest Toxic Unit is correlated best with the ecological effect. Accordingly, we used the maximum toxicity exerted by any of the pesticides identified to link the contamination with $SPEAR_{pesticides}$; the TUmax (Liess and von der Ohe 2005, Schaefer et al. 2007, Knillmann et al. 2018, Liess et al. 2021).

Typically, a multitude of pesticides are entering surface waters simultaneously or sequentially and also act in concert. The approach, to base the assessment of the most toxic substance found in a certain spraying season takes into account all combined effects of pesticides so that the assessment using the $SPEAR_{pesticides}$ approach gives a realistic worst-case assessment also of single substances.

[20] Liess, M. and P. C. von der Ohe (2005): Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry 24(4): 954–965.

[21] Knillmann, S., P. Orlinskiy, O. Kaske, K. Foit and M. Liess (2018): Indication of pesticide effects and recolonization in streams. Science of The Total Environment 630: 1619-1627.

Realistic Consideration of Effect

The use of effect-based indicators is a powerful tool for identifying the toxic pressure of pesticides. The invertebrate-based indicator $SPEAR_{pesticides}$ uses trait information of taxa to identify pesticide pressure and the ecological effects in streams. Due to the use of trait information that are independent of that are independent of species rated vulnerability $SPEAR_{pesticides}$ was successfully applied to indicate pesticide pressure in streams of different geographical regions worldwide including Europe (Liess and von der Ohe 2005, Schaefer et al. 2007, Liess et al. 2008, Münze et al. 2015, Orlinskiy et al. 2015, Münze et al. 2017, Knillmann et al. 2018, Liess et al. 2021), Russia (Beketov and Liess 2008c), Australia (Schaefer et al. 2011), USA (Chiu et al. 2016) and South America (Hunt et al. 2017).

In the environment pesticide effects are not only determined by the toxicity of the pesticide but also by the environmental context. Sensitivity to toxicants is increased by a factor of 10 to 100 according to the magnitude of environmental stress (Liess et al. 2016a). Time for recovery from pesticide effects is strongly depending on the magnitude of interspecific competition (Knillmann et al. 2012b). Sequential exposure can lead to culmination of single effects. Populations exposed over several generations to repeated pulses of low concentrations of pesticides may suffer a multigenerational culmination of low-dose effects (Liess et al. 2013).

Application of Approach

$SPEAR_{pesticides}$ is calculated by using the following equations:

$$SPEAR_{pesticides} = \frac{\sum_{i=1}^{n} \log(4x_i+1) * y}{\sum_{i=1}^{n} \log(4x_i+1)}$$

where n is the total number of taxa in a sample, x i is the abundance of taxon i (given as individuals per m −2), and y is set to 1 if taxon i is classified as “at risk” (Liess and von der Ohe 2005), i.e., vulnerable to pesticides under regular exposure events and set to 0 otherwise. Abundance data were log(4× + 1)-transformed to decrease the influence of populations with mass developments.

$$SPEAR_{pesticides} = \frac{SPEAR_i}{SPEAR_{reference}}$$

where $SPEAR_i$ represents the indicator value of a macroinvertebrate community at a specific site *i* and time point and $SPEAR_{reference}$ represents the mean $SPEAR_{pesticides}$ value under reference conditions regarding toxic pressure.

$SPEAR_{pesticides}$ can be calculated with the species classification of the software “SPEAR Calculator”, http://www.systemecology.eu/spearcalc/index.en.html. Traits applied are (i) the physiological sensitivity according to Wogram and Liess (2001) and Von Der Ohe and Liess (2004), (ii) generation time, (iii) ability to recolonise, (iv) exposure during main period of application. The trait values for a great number of species can be obtained through the software “SPEAR Calculator”. Figure 1 gives an overview on the approach of $SPEAR_{pesticides}$.

2.10.1.2 Status of the Model

$SPEAR_{pesticides}$ was developed by Liess and von der Ohe (2005) and is currently successfully applied to indicate pesticide pressure in streams of different geographical regions worldwide as indicated by numerous peer reviewed publications. The approach is constantly improved with the latest iteration published by Liess et al. (2021). All relevant information to calculate SPEAR are available in the respective publications (Liess and von der Ohe 2005, Liess et al. 2021) and is included with the free software INDICATE (http://www.systemecology.eu/indicate/). From within this program all trait information on species can be obtained.

2.10.2 Model Description

2.10.2.1 Problem Definition

Context in which the model will be used

Pesticides are not allowed to be registered if they fail the first tier of ERA, unless it has been established that under field conditions they exert no unacceptable impact on the viability of exposed species (Council Directive 97/57/EC 1997). For this purpose, a refined risk assessment is performed under more realistic ecological conditions, e. g. in mesocosms to assess the risk for aquatic invertebrates. $SPEAR_{pesticides}$ offers the ultimate level of realism as exposure includes multiple pesticide exposure; effect includes realistic context related to additional environmental stressors and biotic interaction. Accordingly, $SPEAR_{pesticides}$ reflects the field situation.

Specification of the question(s) that should be answered with the model

$SPEAR_{pesticides}$ identifies a realistic assessment factor that can be multiplied on the LC50 identified by the standard toxicity test (*Daphnia magna*, *Chironomus riparius*: whichever is more sensitive).

Specification of necessary model outputs and protection goals

$SPEAR_{pesticides}$ addresses field realistic multiple pesticide exposure and field relevant context sensitivity. Outcomes can directly be transferred into protection goals.

Domain of applicability of the model

Direct pesticide effects on aquatic invertebrates.

Why is the model being used?

To identify field relevant pesticide effects on aquatic invertebrates.

What protection goal is being addressed?

Identification of unacceptable long-term effects on edge of field waterbody invertebrate populations.

What outputs are required?

Identifies a realistic assessment factor that can be multiplied on the LC50 identified by the standard toxicity test.

How was the species chosen?

Species that are occurring in edge of field streams.

Which other species/groups are being covered by the chosen one(s)?

Not necessary as all relevant species are considered.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate?

Field data have been used to successfully validate the model.

2.10.2.2 **Supporting Data**

Summary of the key data used in the model for development and evaluation

Model development was based on a literature review on the ecology of invertebrate species (Wogram and Liess 2001, Von Der Ohe and Liess 2004, Liess and von der Ohe 2005, Knillmann et al. 2018). Information are accessible via INDICATE (http://www.systemecology.eu/indicate/). The most comprehensive dataset can be accessed through (Liess et al. 2021): (Data publication simultaneous to this paper via the data publisher PANGAEA, under embargo and will be publicly available on the 30.09.2022 at https://doi.org/10.1594/PANGAEA.931673. Title: The lowland stream monitoring dataset (KgM, Kleingewässer- Monitoring) 2018, 2019.

Assessment of the quality of the data

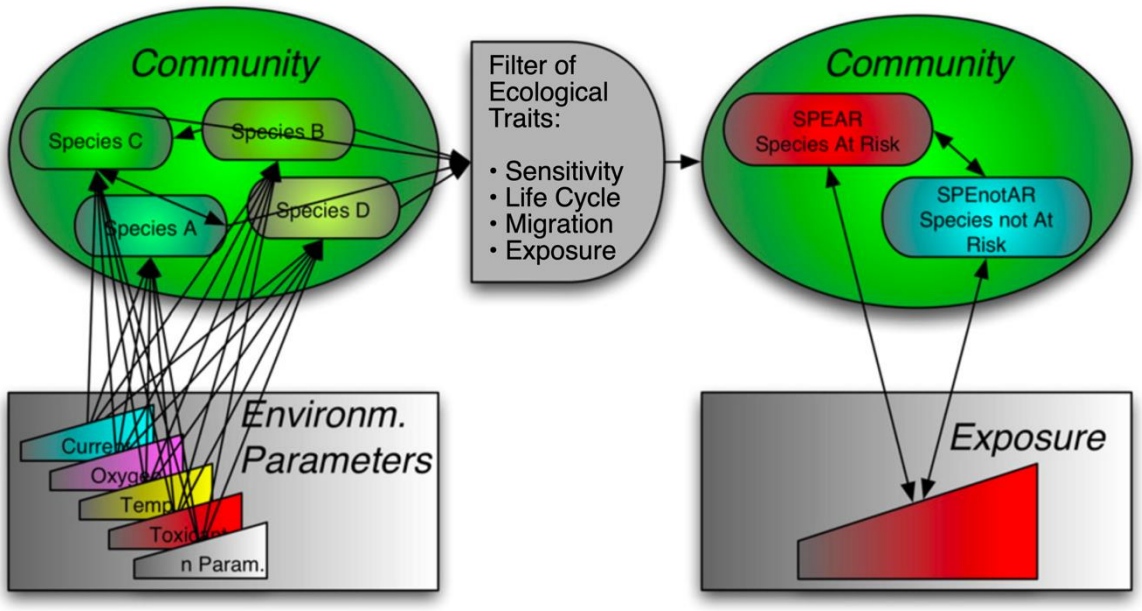
Trait data were validated in field investigations.

2.10.2.3 **Conceptual Model**

Description of the model concepts including a diagram

SPEAR$_{pesticides}$ is a trait-based model identifying traits that are sensitive to agricultural pesticide exposure. Traits applied are (i) the physiological sensitivity according (Wogram and Liess 2001, Von Der Ohe and Liess 2004), (ii) generation time, (iii) ability to recolonise, (iv) exposure during main period of application. The trait values for a great number of species can be obtained through the software "SPEAR Calculator". Figure 1 gives an overview on the approach of SPEAR$_{pesticides}$.

Figure 19: SPEAR – Overview on the Conceptual Model



Graph reproduced from Liess et al. (2008).

Identify the main components and processes in the system

Traits applied are (i) the physiological sensitivity according to Wogram and Liess (2001) and (Von Der Ohe and Liess 2004), (ii) generation time, (iii) ability to recolonise, (iv) exposure during main period of application.

How the effects of the chemicals are modelled

Pesticide exposure changes the trait composition.

How the components and processes are linked

Trait modalities are logically linked with a Boolean operation. If all modalities of a taxon are sensitive, the taxon is defined as vulnerable. Otherwise invulnerable. Then the ratio of vulnerable taxa is calculated and related to the maximum toxic unit of pesticides measured. Detailed description (Liess and von der Ohe 2005).

2.10.2.4 Formal Model

Identification of the model variables

Individuals are described by traits that are known to be related to pesticide vulnerability (Liess and von der Ohe 2005).

Identification of the model parameters

Trait thresholds were identified using field monitoring data.

Description of the most important model equations or algorithms

$SPEAR_{pesticides}$ is calculated by using the following equations:

$$SPEAR_{pesticides} = \frac{\sum_{i=1}^{n} \log(4x_i+1) * y}{\sum_{i=1}^{n} \log(4x_i+1)}$$

where n is the total number of taxa in a sample, x i is the abundance of taxon i (given as individuals per m −2), and y is set to 1 if taxon i is classified as “at risk” (Liess and von der Ohe 2005), i.e., vulnerable to pesticides under regular exposure events and set to 0 otherwise. Abundance data were log(4× + 1)-transformed to decrease the influence of populations with mass developments.

$$SPEAR_{pesticides} = \frac{SPEAR_i}{SPEAR_{reference}}$$

where SPEAR i represents the indicator value of a macroinvertebrate community at a specific site i and time point and SPEAR reference represents the mean SPEAR pesticides under reference conditions regarding toxic pressure.

2.10.2.5 Computer Model

Description of the model implementation

Model is implemented in the package INDICATE (http://www.systemecology.eu/indicate/).

Checking the computer model for errors, bugs and inconsistencies in the code

The code is tested extensively by applying various monitoring studies.

Demonstrate that the computer model performs as indicated by the conceptual and formal models

The model is tested extensively by applying various monitoring studies.

2.10.2.6 The Environmental Scenario

Description of the environmental scenarios, i.e. the environmental context **in which the model is run**

The environmental scenario is based on numerous edge-of-field waterbody monitoring investigations in various geographical regions worldwide. Maximum realism is guaranteed.

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

Field investigation insure for realistic combination of abiotic, biotic and agro-environmental parameters.

2.10.2.7 Parameter Estimation

Description of the model parameter estimation

Trait modalities were defined based on field investigations.

Parameters estimated from the literature — what are the sources and why are these appropriate?

The selection of traits is based on a literature review on the effects of pesticides on aquatic invertebrate communities in streams.

2.10.2.8 Parameters obtained from calibration — how and why this was done?

The selection of traits is based on a literature review on the effects of pesticides on aquatic invertebrate communities in streams.

2.10.2.9 Sensitivity and Uncertainty Analysis

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

Sensitivity and uncertainty analysis for SPEAR is reflected by the R^2 between contamination and SPEAR value. The R^2 is typically above 0.6. Accordingly, more than 60 percent of variation in community composition can be attributed to variance in contamination.

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

The R^2 is typically above 0.6.

2.10.2.10 Comparison with Measurements

Description of comparisons of model output with independent data

Model predictions have been tested with several independent field data.

Demonstration that the model output provides an adequate match to data patterns

Model predictions have been tested with several independent field data including Europe (Liess and von der Ohe 2005, Schaefer et al. 2007, Liess et al. 2008, Münze et al. 2015, Orlinskiy et al. 2015, Münze et al. 2017, Knillmann et al. 2018, Liess et al. 2021), Russia (Beketov and Liess 2008c), Australia (Schaefer et al. 2011), USA (Chiu et al. 2016) and South America (Hunt et al. 2017).

2.10.2.11 Model Use

Explanation of how the model conforms to the requirements set in the problem definition

The model provides a field validated assessment factor on standard test species (*Daphnia magna*, *Chironomus riparius*) It therefore conforms to the problem definition.

Description how the model works (user manual).

A description is provided on the webpage INDICATE (http://www.systemecology.eu/indicate/).

Description of the pesticide parameters values used in the model

The pesticide parameter values do not need to be specified as the field investigation warrants that input, degradation, exposure, uptake and long-term effects are adequately considered.

Description of the specific assessment including a discussion of the most important results

The model provides a field validated assessment factor on standard test species (*Daphnia magna*, *Chironomus riparius*). It therefore conforms to the problem definition. Based on the field results the assessment factor for LOEC is 2.000 (Liess et al. 2021), NOEC 10.000 compared to the more sensitive test species.

2.10.3 Model Evaluation

2.10.3.1 **Problem Definition**

The regulatory context in which the model is run

Pesticides are not allowed to be registered if they fail the first tier of ERA, unless it has been established that under field conditions they exert no unacceptable impact on the viability of exposed species (EU 1997). For this purpose, a refined risk assessment is performed under more realistic exposure and ecological conditions. However, for the exposure temporal variation of concentration and effects of multiple pesticide exposure is not considered appropriately in the current risk assessment. Similar for the prediction of ecological effect major limitations are eminent as respective effects are not considered: (i) long-term effects do not consider transgenerational effects (Pieters and Liess 2006), (ii) multiple stressor greatly increase effects of toxicants (Liess et al. 2013, Liess et al. 2016a) (iii) time for recovery is delayed by the presence of interspecific competitors (Knillmann et al. 2012b) and also depends on the quality and distance of recovery sources (Liess and von der Ohe 2005, Knillmann et al. 2018), (iv) the magnitude of pesticide adaptation is related to the ecological context (Becker and Liess 2015, Becker and Liess 2017). Each of the described processes may alter the identification of a regulatory acceptable concentration by more than a factor 10. Accordingly, the uncertainty of the current risk assessment procedure sums up to several magnitudes.

In the light of this situation the SPEAR concept offers a dramatic improvement by combining a Tier 1 effect estimation with field-based exposure and effect estimations.

The question that has to be answered with the model

The relevant assessment factor.

The available knowledge and data relevant to the risk assessment question

All aspects related to the RA question are available.

The outputs required to answer these questions including performance criteria for the regulatory model

The output required consists of a field-based assessment factor.

The species to be modelled

All species present in edge of field water bodies.

Requirements for the environmental scenarios to be used in the risk assessment

The field investigations need to reflect the typical range of scenarios in agricultural landscapes.

2.10.3.2 **Supporting Data**

Are the data fit for purpose in view of the problem definition?

Yes.

Has the quality of the data used been considered and documented?

Yes, see publications.

Have all available data been used? If not, is there a justification why this information has not been used?

Yes.

2.10.3.3 Conceptual Model

Are the specific protection goals sufficiently well addressed by the model?

Yes.

Are the modelling endpoints relevant to the specific protection goal?

Yes.

Is the modelling approach justified?

Yes.

Is the conceptual model logical?

Yes, according to a broad ecotoxicological consensus the traits used are relevant to predict the ecological effects of agricultural pesticide application.

Are the processes included in the model relevant to the addressed issue?

Yes – see above.

Are the links between different processes to the variables logical?

Yes.

Are the temporal and spatial scales relevant in regard to the problem definition?

Yes, field investigation.

2.10.3.4 Formal Model

Are the most important model assumptions justified by the modeller?

Yes, see publications.

Are the most important mathematical equations described?

All mathematical equations are described.

Is there a description of the variables and parameters including their meaning and unit?

Variables and parameters are clearly described and justified.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

Yes, see publications.

Are references supporting the equations been provided?

Yes, see publications.

2.10.3.5 Computer Model

Is there a comprehensive and transparent description of the computer model?

Yes, see publications.

Is the computer code well readable and is it available?

Yes, see publications.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

Yes, see publications.

2.10.3.6 The Environmental Scenario

Is the scenario representative for the risk assessment under consideration?

Yes, field investigation.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

Yes, see publications.

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

Yes, see publications.

Is the level of conservatism placed into the scenarios appropriate?

Yes, field investigation.

2.10.3.7 Parameter estimation

The model parameter estimation has been adequately documented?

Yes, see publications.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Yes, see publications.

Were the estimated parameter values realistic?

Yes, see publications.

Are the data sources sufficiently documented?

Yes, see publications.

2.10.3.8 Sensitivity and Uncertainty Analysis

Has the sensitivity analysis been adequately documented?

The equivalent of the sensitivity analyses is reflected in the correlation between exposure and effect. Details see publications.

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

Yes.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

Yes, see publications.

Has the uncertainty analysis been adequately documented?

Yes, see publications.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

Yes, see publications.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

Yes, see publications.

Uncertainty is propagated to the model results?

Yes.

Have confidence intervals been estimated and has this information been used in further model use?

Yes, see publications.

2.10.3.9 Comparison with Data from Independent Measurements

Have the performance criteria for the model been predefined in the problem definition?

Yes, see publications.

Are the model outputs that are compared relevant in view of the problem definition?

Yes.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

Yes, see publications.

Is the dataset relevant in view of the problem definition?

Yes.

Is the fit of model output to the data good enough?

Yes, see publications.

Has the performance of the model been reported in an objective and reproducible way?

Yes, see publications.

2.10.3.10 Model Use

Is a user manual available?

See INDICATE: http://www.systemecology.eu/indicate/

Have all aspects of the modelling cycle been documented?

Yes, see publications.

Has a summary sheet been provided by the modeller?

No summary sheet is publicly available.

2.10.3.11 Evaluation of the suitability of the model for regulatory purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

The model is presented on http://www.systemecology.eu/indicate/ together with the authors' contact addresses.

Is a version control system implemented?

Yes, see http://www.systemecology.eu/indicate/.

2.10.3.12 Overall Judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

The application of $SPEAR_{pesticides}$ as an empirical model for the prediction of pesticide effects on freshwater macroinvertebrates in small streams seems highly suitable for regulatory purposes. The approach predicts the propagation of individual-level effects on a standard surrogate species (acute LC50 for *D. magna* or a more sensitive standard test species such as *Chironomus* sp.) to changes in the whole macroinvertebrate community composition. The prediction is deduced from a regression of the abundance of vulnerable macroinvertebrate taxa (quantified as SPEAR value) vs. individual-level toxicity in the water (quantified as toxic unit of the most toxic pesticide, TU_{max}); this regression is based on field observations that are representative for Central Europe in terms of typical exposure scenarios, biotic and abiotic conditions.

This way, the model implicitly covers those processes that are highly relevant but currently not considered in ERA:s (i) established methods do not consider effects of multiple pesticide exposure;: (ii) the assessment of chronic effects does not consider transgenerational effects (Pieters and Liess 2006), (iii) multiple stressors greatly increase effects of toxicants (Liess et al. 2013, Liess et al. 2016a) (iv) time for population recovery is delayed by the presence of interspecific competitors (Knillmann et al. 2012b) and also depends on the quality and distance of source populations for recolonization (Liess and von der Ohe 2005, Knillmann et al. 2018), (v) the magnitude of pesticide adaptation is related to the ecological context (Becker and Liess 2015, Becker and Liess 2017). Each of the described processes may alter the identification of a regulatory acceptable concentration by more than a factor of 10. Accordingly, the uncertainty of the current risk assessment procedure sums up to several orders of magnitudes. As a bioindicator, $SPEAR_{pesticides}$ can be used to validate effect predictions by the existing procedures in ERA. As a model for prospective ERA, $SPEAR_{pesticides}$ implicitly covers the processes mentioned above and can thus decrease uncertainty associated with the propagation of individual-level effects in standard tests to the whole community.

The $SPEAR_{pesticides}$ model has been fully calibrated to observational data, therefore the concept of model validation with independent data is of little help here (see also the discussion on the same issue for GUTS in section 2.2.3.9). Validation with independent data is mainly informative to assess whether the model could be used outside its domain, i. e. under exposure and ecological conditions that differ from those in the Central European streams used for calibration. Applications of $SPEAR_{pesticides}$ in Northern and Southern Europe, Siberia Australia, Argentina and Kenya showed that the concept is working across biogeographic regions, climate zones and farming systems, though the trait database and the established SPEAR vs. TU_{max} relationship should be modified to the local conditions in order to improve the model fit (Schaefer et al. 2011, Schaefer et al. 2012, Hunt et al. 2017, Ganatra et al. 2021). The model thus seems robust against moderate deviations in exposure and ecological conditions from typical Central European conditions which could arise from an unusual application pattern and exposure profile of a pesticide to be assessed.

To assess the accuracy of the $SPEAR_{pesticides}$ model when used within its domain of small Central European streams, it is more informative to assess the goodness of fit of the SPEAR vs. TU_{max} regression. The R^2 of 0.57 (Knillmann et al. 2018) mainly reflects the actual variability in effect propagation due to variable exposure profiles and ecological conditions (e. g. communities) in the field. However, the R^2 additionally reflects uncertainty in the modelling approach due to differences in pesticide properties. E. g., two pesticides may show similar acute toxicity to the reference species *D. magna* but may differ in their toxicity to other macroinvertebrates or in their chronic toxicity, which will affect the potential of effect propagation to the community level. Therefore, $SPEAR_{pesticides}$ may be used as a screening step to predict a range of community effects that can be expected from the predicted environmental concen-

tration (PEC) and the observed individual-level effects (acute LC50) of a pesticide. A refined risk assessment may then evaluate whether community effects will be at the lower margin of this range due to the specific properties of the pesticide.

The SPEAR related protective concentration, the AC_{field} that is available for 22 primarily invertebrate-toxic pesticides identifies an extrapolation factor related to acute LC50 values of about 2000 protecting 95% of streams; a factor exceeding the acute regulatory Tier 1 "assessment factor" (100) by 20. To protect 99% of streams the respective extrapolation factor would amount to 18,000, a log TU_{max} of −4.25. However, the exposure to RAC ratio was found to explain $SPEAR_{pesticides}$ equally well as the exposure to LC50 ratio (R2 =0.44 versus R2 =0.43). This shows that the RAC values are related to the ecological effect as shown in the cause-effect relationship. Nonetheless, their compliance would cause unacceptable effects in 14% of agricultural stream sections; 86% would be protected. To protect 95% or 99% of streams, respectively, the RAC for invertebrate-toxicity driving pesticides required an additional assessment factor of 5.3 or 40.2 (Liess et al. 2021).

Taken together, $SPEAR_{pesticides}$ is a valuable tool to assess the level of protection that is achieved with the current use of surrogate species in ERA for the whole community. The application of SPEAR in prospective ERA, however, would require the establishment of a new SPG that refers to the SPEAR value as a community-level endpoint.

2.10.4 Qualitative Assessment of Uncertainties

2.10.4.1 Potential for Underestimation of Real Risk

- The pesticide effects are identified against the situation in control sites. In case also the control sites are affected by pesticides (through long distance transport of pesticides or long-distance pesticide effects), effects of contaminated sites are underestimated. This effect may not be too relevant as control sites in Sweden with no relevant pesticide effect in their vicinity showed a similar SPEAR value as control sites in closer vicinity to streams with pesticide contamination present (Schaefer et al. 2007).
- It may be that a small proportion of taxa is more sensitive than the majority of vulnerable species. In case this proportion is small it could not sufficiently influence the $SPEAR_{pesticides}$ response. Accordingly, it may be possible that for these taxa an assessment factor of 2.000 is not sufficient.

2.10.4.2 Potential for Overestimation of Real Risk

- As the assessment is based on field data, all effect-determining factors are included. However, the assessment is based on a likelihood that is described in Liess et al. (2021).

2.10.4.3 Potential for Uncertainty in Either Direction

- As the assessment is based on field data all effect determining factor are included. However, the assessment is based on a likelihood that is described in Liess et al. (2021).

2.11 AQUATOX

Evaluation by Jeremias Becker

2.11.1 General Information

2.11.1.1 Background and Concept

Overview

AQUATOX is one of the most complex and probably the best-known mechanistic ecosystem model available for the effects of toxicants (except metals) and other stressors in freshwater systems (Koelmans et al. 2001, Lombardo et al. 2015). The model can be applied to systems of varying complexity, ranging from simple artificial conditions such as beakers in the laboratory to complex stream or lake ecosystems with more than 40 species and numerous abiotic compartments modelled at the same time. AQUATOX is supported by the Environmental Protection Agency of the United States (USEPA) and has a long history of development and application. Here we focus on the scientific publication of the model in Park et al. (2008)[22] and on the technical documentation for AQUATOX release 3.1 plus (Park and Clough 2014)[23] and for release 3.2 (Park and Clough 2018)[24].

In contrast to most population models, but similar to most other ecosystem models, AQUATOX is not individual-based. Taxonomic or functional groups of organisms and different forms of detritus are represented by compartments of biomass that increase and decrease following a set of coupled differential equations. The model therefore does not predict population sizes but (bio)mass, such as mg suspended sediment per L, or g biomass of grazing invertebrates per m^3. By default, simulations are reported in time steps of 1 day (though the differential equations are numerically solved in shorter, variable time steps depending on their stiffness, based on Runge-Kutta methods).

AQUATOX simulates the combined environmental fate and effects of toxic chemicals and conventional pollutants, such as nutrients, organic wastes and inorganic sediments, in various abiotic and biotic compartments. The modelled biotic compartments cover all trophic levels and include phytoplankton, periphyton, submerged macrophytes (incl. bryophytes as special subclass), zooplankton, insect larvae, molluscs and fish. Each of these functional groups is subject to a specific set of biological processes, and can consist of various taxa that are separately parameterized and simulated. Abiotic compartments in the basic model include the water column (which can be separated in hypolimnion and epilimnion in lake scenarios), inorganic sediment, and eight different compartments of detritus at the bottom surface. The detrital forms include dissolved, suspended, sedimented and buried detritus, each in a refractory and a labile (readily decomposed) form. This complexity in abiotic compartments was considered necessary for the realistic simulation of the detrital food web, the biological oxygen demand, and the bioavailability of toxicants (Park et al. 2008). Decomposition and conversion of refractory to labile detritus are modelled as first-order equations with multiplicative limitations for suboptimal environmental conditions, but without explicit simulation of microbes (they are considered part of the detritus). The user can exclude abiotic and biotic compartments and this way reduce complexity in

[22] Park, R. A., J. S. Clough and M. C. Wellman (2008): AQUATOX: Modeling environmental fate and ecological effects in aquatic ecosystems. Ecological Modelling 213(1): 1-15.

[23] Park, R. A. and J. S. Clough (2014): AQUATOX (Release 3.1 plus) - Modelling environmental fate and ecologcial effects in aquatic ecosystems. Volume 2: Technical Documentation. United States Environmental Protection Agency (USEPA). Office of Water (4305). Office of Science and Technology; Washington DC 20460 , USA. Report No. EPA-820-R-14-007.

[24] Park, R. A. and J. S. Clough (2018): AQUATOX (Release 3.2) - Modelling environmental fate and ecologcial effects in aquatic ecosystems. Volume 2: Technical Documentation. United States Environmental Protection Agency (USEPA). Office of Research and Development. Office of Science and Technology; Washington DC 20460, USA. Report No. EPA/600/B-18/241.

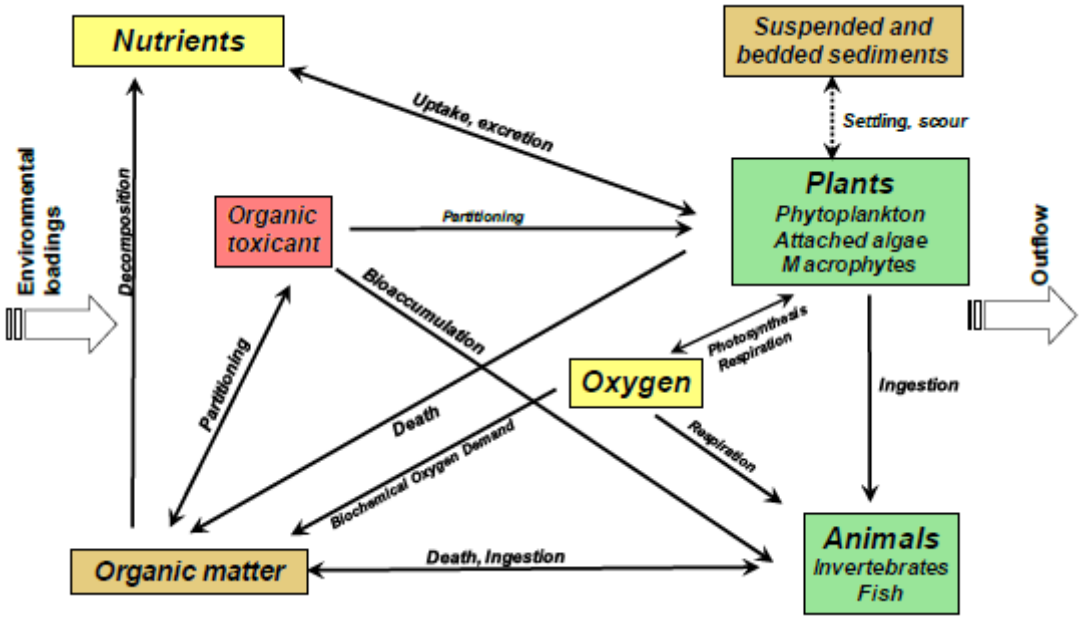
the model. Due to this flexibility, setting a scenario in AQUATOX comprises not only decisions on model parameterization but also on model structure.

AQUATOX models a high number of ecological processes. Similar to other ecosystem models such as CATS (Naito et al. 2003) and CASM (Bartell et al. 1999), algae are subject to loading (immigration) from upstream, photosynthesis, respiration, excretion, non-predatory mortality, grazing and washout (emigration). In AQUATOX, phytoplankton is additionally subject to sinking, and periphyton is subject to sloughing (Park et al. 2008). Macrophytes share many functions with algae in the model. Different features include light extinction by periphyton that contributes to the overall light limitation of macrophytes. Nutrient limitation is only included for the subclass of bryophytes (moss), because macrophytes are considered to obtain their nutrients from bottom sediments that is not included in the simulation. Additionally, macrophyte mortality is not only subject to high but also to low temperatures (die-back in winter) and to breakage at high water velocity. Animals are subject to various physiological and ecological processes considered also in other ecosystem models, such as feeding, assimilation, excretion, respiration and predation. Specific for AQUATOX is the possible reduction of ingestion due to sublethal toxicant effects and suspended sediments, and due to limitations from habitat preferences (Park et al. 2008). Relative food preferences of each taxon are considered, and taxa switch preferences if a preferred prey becomes too rare. The proportion of ingested prey that is actually assimilated varies with prey type, the rest is discarded or defaecated. Unlike most other models, fish mortality in AQUATOX is subject to high levels of ammonia and suspended sediment, and to low levels of oxygen (effects of diel oxygen fluctuations can be optionally simulated). Respiration in the model depends on temperature and, in case of fish, on population density. In AQUATOX, zooplankton and fish avoid an anoxic hypolimnion by migration to the epilimnion. Conversion of refractory to labile detritus further remineralization to dissolved nutrients by microbes is modelled as being limited by nitrogen, suboptimal temperature, pH, and dissolved oxygen. pH and dissolved oxygen are affected by biological activities, and in turn affect the development of biota. This way, the model simulates various feedback loops; e. g., excessive nutrients can cause an algal bloom followed by increased decomposition that will decrease the concentration of dissolved oxygen and lead to fish kill.

AQUATOX can simulate the fate and effects of up to 20 organic toxicants at the same time, but is not considered fit for metals (Park et al. 2008). The built-in fate module includes partitioning of toxicants among organisms, suspended and sedimented detritus, suspended and sedimented inorganic sediments, and water. Aqueous exposure by contact with the water column is modelled using a 1-compartment toxicokinetic (TK) module for each taxon. Additionally, fate processes include prey consumption and egestion, volatilization, hydrolysis, photolysis, ionization, and microbial degradation (Park and Clough 2014). Therefore, both aqueous and dietary exposure are covered. The user can specify degradation pathways leading to degradation products that can have their own toxicity. Thus, the fate part describes the environmental fate of toxicants including bioaccumulation within a water ecosystem, but requires external information on the toxicant input (loading) to the system. For this purpose, AQUATOX may be coupled with the watershed loading model HSPF (Hydrologic Simulation Program Fortran).

The built-in ecotoxicology module can simulate direct lethal and sublethal effects of the modelled toxicants and of their degradation products in each of the modelled taxa. Additionally, various indirect effects can emerge from the simulation, such as the release from interspecific competition and predation, recycling of nutrients and of persistent toxicants from killed organisms, and the loss of food base for animals (Park and Clough 2014).

Figure 20: AQUATOX – Simplified Conceptual Model



Simplified overview on the model concept of AQUATOX. Graph reproduced from Park and Clough (2018).

Ecotoxicology Module

AQUATOX relates direct effects of toxicants to their internal concentration within organisms, i. e. within a biotic compartment. Direct effects are calculated separately for each taxon and each toxicant. Effects of multiple toxicants (and of other environmental stress) on the same taxon are thus considered to be additive (effect addition).

The ecotoxicology module for direct effects was designed to require only minimum ecotoxicological information as input, i. e. the acute LC50 of a given toxicant for a given species as observed after a given time of constant exposure in a standard acute test. Otherwise, AQUATOX would be hardly applicable due to the limited availability of detailed ecotoxicological information on toxicants in all the species of a modelled ecosystem. Nevertheless, the modelled direct effects vary dynamically with the exposure profile. This is done with an extended dose-response approach that scales the cumulative effects observed by the end of the standard test to the concentrations and exposure time experienced in the simulation (see details in sections 2.11.2.3 and 2.11.2.4): In short, first the provided external LC50 is converted to an internal LC50 that is considered to exponentially decrease with exposure time. Therefore, the sensitivity of organisms in AQUATOX increases with exposure time, and consequently direct effects of a given internal concentration become more severe with prolonged or repeated exposure. Each time step, this time-dependent internal LC50 is used in a dose-response equation to calculate the cumulative mortality that would have been reached if the current internal concentration was constant for the whole time of previous exposure (over one generation time, up to one year). The actually experienced mortality in the time step is then calculated as the increment in cumulative mortality calculated for the current as compared to the previous time step.

A fraction of the biomass of each invertebrate and fish species is considered to be gametes. Gametes are modelled as more sensitive to many stressors and toxicants than the remaining fraction of biomass of the same species. The fraction of gametes changes when the population approaches carrying capacity due to stress from intraspecific competition, but can also be changed when the user specifies specific spawning dates.

In addition to acute mortality (modelled as decrease in biomass), AQUATOX can simulate the following sublethal effects: reduced photosynthesis, accelerated sinking of phytoplankton, increased sloughing (and washout) of periphyton, increased drift (and washout) of invertebrates, reduced growth (in animals split into reduced food consumption and reduced assimilation of consumed food), and reduced reproduction. The strength of sublethal effects is derived from the calculated mortality using application factors, i. e. sublethal:lethal ratios that relate the various EC50 for sublethal effects to the LC50 for mortality. Application factors that have been established for one toxicant and taxon can be applied also to other taxa assuming they are similar, so that AQUATOX requires at minimum LC50 values for all taxa and additionally an EC50 for a single taxon in order to simulate sublethal effects.

The ecotoxicology module requires a whole set of assumptions that is an important source of uncertainty for the modelling results (see discussion in sections 2.11.3.3 and 2.11.3.4). AQUATOX represents a toxicokinetic (TK) but not a real toxicodynamic (TD) model, particularly because it relates effects directly to the internal concentration and not to a budget for damage or energy like in the various variants of GUTS and DEB. Therefore, there is no delay in the onset of effects after internal exposure has started, and the affected process rates of a taxon (but not its biomass!) recover instantaneously from sublethal effects when internal exposure has ended.

Additional Optional Processes

AQUATOX can simulate scenarios with a number of separate but linked ecosystem segments, such as riffles and pools in a stream, or river segments upstream and downstream of a point source of pollution (Park and Clough 2014). This way, it is possible to introduce a spatial component and simple meta-population dynamics in AQUATOX: The fraction of phytoplankton, zooplankton and zoobenthos that is subject to washout in an upstream section is moving to the next downstream section. The amount of organisms that immigrates from non-modelled upstream sections, i. e. the loading of the simulation with new biomass, is estimated from the user-specified upstream length of the stream. This approach is similar to the HSPF model that can be coupled to AQUATOX for the loading of toxicants (Park et al. 2008). In AQUATOX, recolonization in streams results therefore from an equilibrium of biomass loadings from inflow and of the velocity-dependent outflow of biomass. The user can specify seasonal migration of taxa between segments by setting dates and the proportion of migrating biomass, e. g. to simulate seasonal vertical migration of fish in a lake (in addition to the avoidance of an anoxic hypolimnion mentioned above). AQUATOX offers also a migration module for anadromous fish species that spawn in freshwater but leave the modelled ecosystem for most of the time to feed in saltwater. Additionally, for each taxon, a fixed rate of loss from fishing or from predation by organisms outside the simulated water body can be set.

Fish, and since Release 3.2 also oysters and crabs (USEPA 2018), can be simulated as multiple compartments to distinguish different size or age classes. This way, demographic effects or bioaccumulation in larger individuals can be modelled. Fish can be modelled using multiple age-classes, where the biomass (and body burden) from each age class is promoted to the next age class on the first spawning date each year. Alternatively, fish, oysters and crabs can be differentiated in two size classes with continuous promotion of biomass from one size class to the next. Promotion to the next size class in the model is determined by the rate of organism growth. However, AQUATOX simulates the overall change in biomass of a species without differentiating in growth and reproduction of individual organisms. The growth rate of organisms is thus estimated in the model as the difference between gain in biomass from consumption (not from migration) and loss of biomass from all causes other than mortality and emigration (i. e. from defecation, respiration, excretion and gamete production). A fraction of this growth is considered as promotion to the next size class (Park and Clough 2014).

Similarly, the organism growth rate is used to calculate the loss of biomass in insects due to the emergence of aquatic larvae as non-simulated flying or terrestrial adults. Insects often show synchronised emergence in the field. AQUATOX simulates this behaviour using the seasonal temperature cycle.

When the temperature is within a specific range, the promotion rate of insects is doubled. Depending on the selected temperature range, this simulates the emergence of multivoltine species in spring and autumn or of univoltine species in summer.

AQUATOX models only species that feed in the water phase. However, flying emerged insects can provide an important source of food for fishes. Adults as food source may be modelled with a workaround by setting the main metabolic rates (mortality, consumption, respiration etc.) of adult insects to zero and by adding a constant loading of their biomass from inflow (Lombardo et al. 2015). AQUATOX may be also used to assess the toxicant exposure of terrestrial or aerial vertebrates such as shorebirds that feed from the contaminated modelled freshwater ecosystem, but are not explicitly simulated. For this task, terrestrial or aerial vertebrates can be incorporated as a non-dynamic, post-processed variable that collects the amount of toxicant in the prey, assuming that the vertebrates feed solely from the modelled ecosystem. This way, terrestrial and aerial vertebrates do not affect the biomass in the ecosystem through predation, which may be simulated through increased outflow rates (Park and Clough 2014).

A large number of additional options are available. E. g., the user can decide whether effects of nutrient limitation on primary producers should be considered based on external or internal nutrient concentrations. Similarly, the calculation of toxicants can be based on internal or external concentrations (see section 2.11.2.3), and the concentration of freely dissolved toxicants can be modelled as constant, regardless of the simulated uptake by organisms. The default reporting step size can be changed from days to hours. Nitrogen fixation in the ecosystem may be simulated based on the N to inorganic P ratio. An optional sediment diagenesis module adds a simulation of the sediment bed to improve the modelled nutrient fluxes by the decomposition of organic matter in pore-water and the release of nutrients from pore-water to the water column.

Scenarios and parameterization

The parameterization of AQUATOX is fully customizable and allows the user to simulate a large variety of freshwater scenarios. However, full parameterization of AQUATOX is very complex and requires large data sets on numerous abiotic and biotic processes. Therefore, AQUATOX offers five built-in libraries for parameter values concerning plants (growth and physiology), animals (feeding preferences, growth and physiology), chemicals (fate and effects), sites (driving variables and constants such as depth, latitude, light, wind and temperature) and remineralization (e. g. degradation and denitrification rates and their dependencies on environmental conditions such as temperature and pH). All libraries are thoroughly referenced.

The user can choose either to start with one of 30 built-in scenarios from previous studies available (version 3.1 Plus), or to create a new scenario from scratch with the help of the libraries. The built-in studies were performed on different types of ecosystems, mainly from (but not limited to) North America.

Tools for Model Analysis

After the specification of a scenario, the user can run the model in a deterministic or in one of two probabilistic modes (Clough 2014). In the deterministic mode, AQUATOX is run once with and without toxicant loadings (treatment and control run), and all parameters are represented as point estimates (fixed values). A number of tools are available to compare the results of treatment and control runs, including plots of the changes in all desired state variables over time, and plots showing the differences between both model runs. Additionally, a set of similarity indices such as the Steinhaus index, and a large number of biotic and abiotic metrics such as % EPT and Net Primary Productivity can be plotted, either as a dynamic function of time or as an average over a specified time period. Since AQUATOX simulates no individuals, the biotic indices are based on the proportion of biomass, not on the

proportion of individuals (Park and Clough 2014). The two probabilistic modes offer an automated or a manual sensitivity analysis and an automated uncertainty analysis (see section 0).

To facilitate model validation, AQUATOX converts information from some state variables into end-points that are better accessible to measurements in the field. E. g., the model converts algal biomass to chlorophyll a values and computes Secchi depth from the overall extinction coefficient for comparison with optical measurements. As another example, the sediment oxygen demand is calculated by taking the sum of detrital decomposition which is then multiplied with the ratio of oxygen to organic matter.

2.11.1.2 Status of the Model

Development

AQUATOX is the latest in a long series of models, starting with the aquatic ecosystem model CLEAN, which was subsequently improved in consultation with researchers at European hydrobiological laboratories, resulting in the CLEANER series and LAKETRACE (Park et al. 2008, Park and Clough 2014). The MACROPHYTE model, developed for the U.S. Army Corps of Engineers, provided additional capability for representing submersed aquatic vegetation. Another series started with the toxic fate model PEST, developed to complement CLEANER, and continued with the TOXTRACE model and the spreadsheet equilibrium fugacity PART model (Park and Clough 2014).

AQUATOX combined algorithms from these models with ecotoxicological constructs to achieve a truly integrative fate and effects model. The model was then restructured and linked to Microsoft Windows interfaces to provide greater flexibility, capacity for additional compartments, and user friendliness. AQUATOX Release 3 was the result of an effort to combine all of the various versions of AQUATOX into a single consolidated version (Park and Clough 2014). Models that were combined to produce Release 3 included: AQUATOX Multi-Segment version, AQUATOX Estuarine Version, and AQUATOX PFA Model (Perfloroalkylated Surfactants). AQUATOX 3.1 is an update with the addition of constructs for sublethal effects and uncertainty analyses (Park and Clough 2014). In September 2018, the latest release AQUATOX 3.2 was published that improved software compatibility for model input and output and extended the model to the simulation of oyster reefs and marsh-edge environments (USEPA 2018). The development of AQUATOX has been partially funded by the Risk Assessment Division (RAD) of the USEPA Office of Pollution Prevention and Toxics.

A downloadable stand-alone version of AQUATOX for MS Windows is available for free at the USEPA website https://www.epa.gov/exposure-assessment-models/aquatox-31-download-page#download. The website provides also comprehensive information including a user's manual, a quick start guide for model setup and application, a technical documentation, and some technical notes dealing with data requirements, procedures for the appropriate modelling of water flows, and with an example data set that is provided for use as a starting point. Finally, the website provides an overview on model reviews and applications in the scientific literature, and lists sensitivity analyses and validation studies performed by the USEPA.

AQUATOX Release 2 and Release 3 have been subject to a formal peer review by an external panel in 2003 and 2008, respectively, which considered the model as the first reasonable interface for scientists to explore ecosystem level effects from multiple stressors over time. Additionally, the model has been reviewed favourably in at least 12 articles and books (https://www.epa.gov/ceam/peer-review-aquatox).

Validation

Park and Clough (2014) present a number of graphs showing reasonable accordance of model predictions with observed data. The tested predictions include PCB congener bioaccumulation factors (BAF's) for trout in Lake Ontario; biomass of chironomid larvae in chlorpyrifos-exposed ponds in Duluth, Minesota; benthic chlorophyll *a* in Cahaba River, Alabama; chlorophyll *a* in Lake Onondaga, NY; PCB congener BAF's in Lake Ontario trout; and PCB concentrations in selected animals from New Bedford Harbor, Massachusetts. The authors also report a reasonable fit after simultaneous calibration to multiple independent data sets from three Minnesota rivers. This study is not a typical validation in the sense of EFSA PPR (2014b), but may nevertheless increase confidence in the credibility of the model because a good simultaneous fit to several independent data sets shows that the driving process have been captured (see pattern-oreinted modeling, e. g. Grimm and Railsback 2012). The studies presented by Park and Clough (2014) were used to generate comprehensive validation reports available at the AQUATOX website. In general, the model was able to predict abiotic processes well, such as changes in water levels, dissolved oxygen and bioaccumulation factors (BAFs) well. Higher uncertainty was observed for biotic predictions such as for the dynamics in biomass or chlorophyll *a*. Few additional validation studies can be found in the scientific literature. E. g., Sourisseau et al. (2008) parameterized the model to artificial streams and validated its ability to predict the dynamics of biotic and abiotic compartments under conditions without toxicants (see section 2.11.2.9).

However, we found no report that validates the prediction of toxicant effects on any output variable (e. g. a comparison of the predicted and the observed differences in biomass of a species between a control and an exposed scenario). Therefore, the capability of AQUATOX for the most relevant predictions for the ERA of plant protection products has not been validated yet.

Application

AQUATOX has been applied to experimental tanks, ponds and pond enclosures, streams, small rivers, linked river segments, lakes, reservoirs, linked reservoir segments and estuaries, oyster reefs and the marsh-edge environment (Park and Clough 2014, Park and Clough 2018). Overall, the USEPA website lists 68 case studies and 25 reviews dealing with AQUATOX that have been published since the year 2000 (https://www.epa.gov/ceam/selected-publications-aquatox, accessed 31.12. 2020).

In particular, the model has been applied to retrospectively evaluate the ecological impacts of different toxicants in rivers, lakes and estuaries. In rivers, e.g., AQUATOX was used to understand observed effects of PCBs (Preziosi and Pastorok 2008, Rashleigh et al. 2009), trichloroethylene (McKnight et al. 2010) and nitrobenzene (Lei et al. 2008). Zhang et al. (2013) used AQUATIOX to model PCB pollution in Lake Baiyangdian in China. Following site-specific calibration of several model parameters, the authors reported a very high fit of model predictions and observed data for the change of biomass of *Cryptomonas*, diatoms, green algae and blue-green algae over the time of one year. Though the authors named this a validation study, it should not be considered as such, because data used for the calibration and testing were not independent from each other. Lombardo et al. (2015) modelled effects of the anionic surfactant linear alkylbenzene sulfonate and the biocide triclosan in River Thames. Calibration was necessary because using the unmodified literature data resulted in an excessive deposition of labile detritus and the collapse of some aquatic invertebrates after three years of simulation. The authors concluded that food webs with generalist fish species feeding on low trophic levels may be more tolerant to the effects of toxicants as compared to more vertical trophic networks including predatory fish. Both direct and indirect effects could result in biomass responses of similar magnitudes, and indirect effects could both exacerbate or compensate for direct toxic effects in the model.

However, we found no reported of application of AQUATOX in prospective risk assessment in the open literature. Galic et al. (2010)concluded that the use of AQUATOX in ecological risk assessment is lim-

ited by concerns related to unacceptably high uncertainty associated with various key model assumptions and input parameters. This refers particularly to assumptions in the ecotoxicology module (see section 2.11.2.3), though AQUATOX has been also coupled on trial with a more conventional TKTD module. Additionally, the applicability of AQUATOX to model charged chemicals has been questioned (Lombardo et al. 2015). The equations implemented to correct inter-media equilibrium partition coefficients of ionisable organic chemicals consider only the contribution of the unionized fraction to partitioning. However, the assumption that organic ions do not partition to solids is inappropriate for weak organic bases or for fully ionized substances (Franco and Trapp 2010, Lombardo et al. 2015). Moreover, in case of missing data, AQUATOX offers to estimate parameters that determine the internal toxicant concentrations (such as uptake and depuration rates) based on the octanol–water partition coefficients of a toxicant. As this estimation is based on correlations observed from a limited number of neutral organic chemicals, the applicability of the estimates to ionisable organics can result in additional model uncertainty (Park et al. 2008). Similarly, AQUATOX offers to extrapolate ecotoxicological data from related species using the ICE regression models, but the uncertainty associated with these estimates can be substantial (Hickey and Craig 2012).

Additionally, Lombardo et al. (2015) explained the limited applicability in prospective risk assessment with the following reasons: lack of field data for the parameterization and validation of complex food webs; the scarcity of non-standard species tested in ecotoxicological experiments; a lack of agreed environmental scenarios; the need to test the model applicability to a wider range of chemicals; and the need to carry out more comprehensive uncertainty analyses. A particular challenge is the calibration of the food web which is often considered non-feasible given practical limitations (Preziosi and Pastorok 2008). However, AQUATOX offers an automatic calibration procedure to support the parameterization of those parameters that are not accessible to empirical observations.

2.11.2 Model Description

2.11.2.1 Problem definition

Context in which the model will be used

AQUATOX has been developed to simulate effects of organic pollutants in freshwater ecosystems. So far, the model has been mainly used by authorities such as USEPA to retrospectively assess effects of toxicants observed in real ecosystems and to identify suitable measures of risk management and ecosystem restauration. However, this does not preclude future use in prospective ERA of pesticides.

Specification of the question(s) that should be answered with the model

AQUATOX is in principle capable of addressing various questions concerning the effects of toxicants on aquatic populations, communities and ecosystems. These include, but are not limited to, the extent of biomagnification in aquatic species, short- and long-term direct effects of lethal and sublethal concentrations, indirect effects of pesticides via trophic cascades, recovery times of species and communities after pesticide exposure, and consequences on ecosystem functions such as reduction in water quality and food sources for fish-eating birds.

Specification of necessary model outputs and protection goals

The specific protection goals and the model output required to address them vary with the specific model applications. AQUATOX can be applied e. g. to identify exposure regimes that will result in less than a given percent reduction in biomass of any modelled species and thus to predict no observable adverse effect concentrations (NOAEC). The model can be also applied to assess recovery times of affected populations in a realistic environmental context.

Domain of applicability of the model

AQUATOX can be set up for various aquatic ecosystems from artificial test systems to ponds, lakes, reservoirs, streams, rivers, estuaries and marshlands under a broad range of environmental conditions. Built-in scenarios represent mainly temperate and subtropical water bodies in North America. The model can simulate effects of various organic toxicants and of sediment pollution (but not metals) on primary producers, invertebrates and fish, but not on decomposers. Because AQUATOX is not individual-based, demographic effects cannot be considered (except for fish and clams that can be differentiated in size or age classes in the model).

Why is the model being used?

Park and Clough (2014) conclude that the most widespread use of AQUATOX is as a screening-level model to assess the fate and effects of pesticides and industrial organic chemicals in representative environments (ponds and pond enclosures, experimental streams, and a representative estuary). Additionally, the model has been parameterized to a number of field studies to develop water quality targets for nutrients. The model can also provide insights into the impact of invasive species and the possible effects of control measures, such as pesticide application. So far, AQUATOX has not been applied in prospective governmental risk assessment.

What protection goal is being addressed?

This issue is difficult to be addressed out of a specific model application. In principle, AQUATOX can be applied to identify exposure regimes that will result in less than x % short-term or long-term reduction in any modelled species in a realistic environmental context. It can be thus used to address the Ecological Threshold Option according to the EFSA aquatic guidance document for regulatory risk assessment (EFSA PPR 2013). AQUATOX can also be applied to assess recovery time under a given exposure regime and thus to address the Ecological Recovery Option (EFSA PPR 2013).

What outputs are required?

The required output depends on a specific model application. AQUATOX is capable of predicting risk graphs that show the probability of different effect sizes to occur. Such predictions can be plotted for numerous output variables, including the population size of all species modelled, internal concentrations in all modelled biotic and abiotic compartments, various diversity indices, and abiotic parameters describing the water quality.

How was the species chosen?

This item depends on a specific model application. AQUATOX is capable of simulating more than 40 species at the same time. Up to 26 plant species (six diatoms, six green algae, six blue-green algae, two other algae and six macrophytes), 14 invertebrate species (two shredders, two sediment feeders, two suspension feeders, two clams, two grazers, two snails and two predatory invertebrates), and several fish species (divided in forage fish, bottom fish and game fish) can be modelled. The model provides 36 built-in case studies which use different sets of species. Generally, the case studies selected representative species of the above-mentioned groups that were common at the study site and for which sufficient knowledge was available.

Which other species/groups are being covered by the chosen one(s)?

With diatoms, green algae, blue-green algae, other algae, macrophytes, shredders, sediment feeders, suspension feeders, clams, grazers, snails, predatory invertebrates and fish, AQUATOX covers the most important taxonomic groups and guilds at the producer and consumer level, whereas decomposers (bacteria and fungi) are not represented. Decomposition is calculated through an abstract decomposition rate.

What data will be used to evaluate the model and degree of match to patterns required to be judged adequate

This item depends on a specific model application in a dossier.

2.11.2.2 Supporting data

Summary of the key data used in the model for development and evaluation

Calibration of AQUATOX to an ecosystem from scratch is very complex but has been done in a number of case studies, mainly in North America (Park and Clough 2014). The built-in case studies can be loaded and subsequently modified by the user, or new scenarios can be built. For the latter case, the parameters concerning plants, animals, site characteristics, remineralization processes, and toxicants have been arranged in built-in libraries, which can be used if no user-specific information is available.

Additionally, AQUATOX is linked to the webICE (Interspecies Correlation Estimation) database (Raimondo, Vivian et al. 2010). This way, LC50 values for a specific species and toxicant can be extrapolated from known LC50 values for related species or toxicants based on simple regressions.

Assessment of the quality of the data

The quality of the data obtained in the case studies vary. Case studies on natural water bodies comprised extensive research on abiotic and biotic compartments in the model, including gut content analyses to assess food preferences of several species. However, abiotic processes and compartments were generally much better studied than most biotic processes, such as species interactions and the interactive effects of multiple stressors at the individual level. The webICE database uses LC50 values from acute tests that were based on or closely related to the standard procedures. Additional information from these tests (shape of dose-response curves, relation of lethal and sublethal effects) are not considered but assumed to be fixed across different toxicants and species.

2.11.2.3 Conceptual model

Description of the model concepts including a diagram

AQUATOX does not model individuals, but biomass, expressed as ash-free dry weight (Park and Clough 2014). Biomass is distributed to several biotic compartments that represent different species or species groups (that are called taxa here) and by default to eight forms of suspended and sedimented detritus (see below). Additional abiotic compartments include the inorganic sediment bed and the water column that contains suspended inorganic matter and the suspended nutrients N, P and C, with N being separated into nitrate and ammonium. The sediment bed can be optionally simulated in more detail using additional compartments. Up to 20 organic toxicants and their degradation products can be distributed to the modelled compartments. All compartments together make up a modelled site. The compartments are connected by a flexible food web that allows the user to add or delete individual compartments and to change links at every trophic level. For a short overview see section 2.11.1.1 above.

A modelled site is characterized by a set of abiotic conditions and properties such as the ecosystem type (pond, lake, stream, reservoir, enclosure, estuary and marine) that activates or de-activates the simulation of certain processes. Other site properties include the volume, depth, stream velocity, temperature, light, wind, loadings and washout of nutrients, suspended sediment, toxicants and biomass. Volume of the modelled water column can be kept constant, subject to daily user input, or calculated based on inflow, outflow an evaporation. Loadings can be controlled by the external models such as HSPF and can vary over time. Washout depends on stream velocity. Light and temperature change seasonally and can be provided externally or calculated internally from the user-specified latitude and amount of shading by surrounding terrestrial vegetation. The site properties affect various abiotic and biotic processes such as reaeration, toxicant degradation, photosynthesis, periphyton sloughing, invertebrate drift and growth of biomass. Sites in riverine ecosystems are subdivided in three habitats: riffles, runs and pools. The habitats are not modelled explicitly but as fractions of a whole site that can be set by the user. Habitats differ in stream velocity that affects reaeration (see below). The percentage of a taxon that is considered to live within a habitat depends on the fraction of this habitat within a site, and on the relative preference of the taxon for that habitat. Each taxon is exposed to a weighted average water velocity depending on its location within the three habitats. Additionally, when the preference of a taxon for given habitat is zero, presence of this habitat reduces the consumption and availability of light for this taxon. Optionally, a site can be modelled as multiple segments that are modelled separately but are linked via inflow / outflow and migration. This way, heterogeneous parts of an

aquatic ecosystem can be differentiated, such as the epilimnion and hypolimnion in a lake, and non-contaminated upstream and contaminated downstream reaches in a stream.

Biomass and toxicants can pass continuously from one compartment to another through various biotic and abiotic processes that are represented by flux rates and depend on various environmental conditions (see below). As a consequence, responses are simulated as averages for an entire taxon or abiotic compartment; there is no distinction between large effects in a small fraction of individuals and smaller effects in a larger fraction of individuals. Similarly, growth of biomass is principally not differentiated in an increase of individual numbers (abundance) and of individual body mass. Assimilation converts nutrients in the water column to biomass of primary producers, and remineralization converts detrital, respired and excreted biomass back to nutrients in order to close the cycle and maintain mass balance.

Taxa are assigned to taxonomic groups (algae) and guilds (animals) that indicate their position in the food web and serve the calculation of biotic indices (such as % EPT) to support model output analysis. A maximum of 20 algae (6 green, 6 blue-green, 6 diatoms. 2 other), 6 macrophytes, and 37 animal taxa can be modelled (6 suspension feeders, 3 deposit feeders, 4 clams, 2 snails, 2 small predatory invertebrates, 4 predatory invertebrates, and 2 small forage fish, 2 large forage fish, 2 small bottom fish, 2 large bottom fish, 4 small game fish, 4 large game fish. 2 oyster species (special type of the clams) can be modelled with separate size classes for spats, veligers (larvae) and adults. Each fish taxon can be modelled using two age classes (adult and juvenile), and one fish species can be modelled with up to 15 age classes that span multiple years. Decomposers (bacteria and fungi) are not explicitly simulated; their activity is modelled by process rates for the transition between different forms of detritus and dissolved nutrients that depend on the abiotic conditions in the water column. The food web can be further specified by setting individual feeding preferences for each taxon. When the preferred food source is limited, taxa can switch prey preferences.

Taxa are further assigned to one of 11 functional groups that are subject to a specific set of biological processes in the model: In addition to processes shared by all plants, phytoplankton is limited by nutrients and light and is subject to sinking and washout. Periphyton is limited by nutrients, low current (to model insufficient nutrient replenishment and removal of senescent biomass) and by shading from phytoplankton, and is subject to sloughing (increases with current and senescence). Benthic and rooted-floating macrophytes are not limited by nutrients because they are assumed to take up nutrients from the ground below the modelled sediment layer that is outside the model domain. However, benthic macrophytes are also limited by shading. Free-floating macrophytes and bryophytes are limited by nutrients, and bryophytes are also limited by shading, whereas free-floating macrophytes are subject to washout. All macrophytes are subject to breakage at high current. In addition to all processes shared by animals in the model (e. g. respiration, limitation by pH), plankton invertebrates are subject to washout. Necton and benthic invertebrates are subject to drift and scour entrainment; oysters or predatory invertebrates can be optionally modelled using separate size classes and are then subject to promotion / recruitment). Benthic insects are subject to emergence (loss of biomass to flying adults outside the model domain). Fish is subject to scour entrainment and may be represented by age classes that are connected by promotion.

Biomass in AQUATOX is formed by primary producers through photosynthesis that requires assimilation of the modelled dissolved nutrients N, P and C from the water column. Each of the nutrients can limit photosynthesis and the associated primary production. As an exception, benthic and rooted macrophytes form biomass without the need of nutrient assimilation (see above) and thus serve as a source of biomass / nutrients. Excretion converts a fraction of biomass from all taxa to detritus, and a second fraction of biomass directly back to nutrients in the water column. Respiration converts biomass directly to dissolved C. Biomass is transferred to higher trophic levels via feeding / predation and subsequent assimilation. The non-assimilated fraction of ingested biomass is defecated and, together with biomass lost by non-predatory mortality, is partitioned to six different forms of detritus. These

include dissolved organic matter in the water, suspended particulate matter and sedimented matter, each in a labile (degradable) and in a refractory form. The allocation of biomass to the different detrital forms varies with the involved processes of detritus formation and functional groups of organisms: E. g., deceived biomass is generally partitioned to suspended and dissolved detritus; however, deceived bryophytes form only refractory forms of detritus, whereas other organisms contribute mainly to the labile forms. In contrast, excretion contributes only to dissolved detritus (mainly to the labile form). Detritivores feed on suspended and sedimented detritus but assimilate only their labile forms. Microbial colonization (not explicitly modelled) converts refractory forms of detritus to their labile counterparts based on a simple process rate. As an exception, refractory dissolved detritus is converted to labile suspended detritus to consider the accumulation of microbes. Finally, decomposition (also simulated only implicitly by a process rate) re-mineralises biomass in the labile forms of detritus back to dissolved nutrients in the water. Sedimentation of suspended particulate detritus and resuspension of sedimented detritus are modelled as an equilibrium depending on discharge in the basic model. If the amount of one of the sedimented forms of detritus exceeds their initial conditions in the simulation, the excess becomes buried in the sediment bed. Buried labile and refractory detritus act as sinks for biomass, but can be reversed when sedimented detritus drops below the initial conditions. Several optional extensions are included in the software package that add an advanced simulation of inorganic sediment transport (sand-silt-clay module), of the inorganic sediment bed (multi-layer sediment module), and of nutrient fluxes in the sediment bed (sediment diagenesis module) to the basic model.

AQUATOX models N in the water column as nitrate and ammonia (the short-living nitrite is not modelled); both forms are assimilated by primary producers. Ammonia forms from decomposition and excretion. A fraction of ammonium (depending on pH and temperature) is considered to be NH3 that is more toxic to invertebrate and fish than the ionized form of ammonia. The toxicity of both forms is considered additive and is modelled similar to those of other toxicants, but effects are based on external concentrations in the water column. Ammonia is converted to nitrate through nitrification, which is lost from the modelled domain through denitrification. Nitrification and denitrification depend on temperature, oxygen and pH. Fresh nitrate is generated from fixation by blue-green algae; additionally, rooted plants can generate biomass without nutrient uptake (see above) that is then be excreted and decomposed to fresh N. P is only considered present as phosphate in the water column. It is formed from excretion and detrital decomposition, and is assimilated by primary producers. C is considered present as carbon dioxide in the water column. It is formed from respiration and detrital decomposition, and is assimilated by primary producers. Carbon dioxide is also created or lost to the model domain through atmospheric exchange. This process is driven by the existing amount of carbon dioixide in the water, as well as wind, stream velocity and depth.

Mass balance is realized in AQUATOX by tracking the nutrients N and P in the food web. Primary producers take up N, P and C from the water and convert them to biomass using taxon-specific conversion factors that describe the fraction of each nutrient in the photosynthate (the factor for C is constant across all taxa). The fractions of N and P in the biomass are considered to be constant for a taxon but can vary between taxa; they can be set as taxon-specific properties. This implicitly sets also the fraction of C, assuming that the fraction of additional non-modelled nutrients remains constant across compartments. When herbivores feed on primary producers, the assimilated amount of prey biomass is converted to predator biomass which is typically characterized by higher N and P contents. Because the overall amount of biomass has not changed, this process could generate N and P from nothing. As a compensation, AQUATOX eliminates the respective amount of N and C from the water. At the same time, however, the defecated amount of prey biomass is converted to detritus that is characterized by lower N and P contents. Similarly, excretion converts predator biomass to N and P poor detritus, so that nutrients could get lost. Additionally, biomass is lost through respiration and its C (but not N and P) content is released to the water. In these three processes, AQUATOX adds the difference in N and P

to the water, so that all processes balance each other for the most part. A remaining nutrient imbalance that results from assimilation, defecation, excretion and respiration is compensated by the addition of the missing amount of N and P to the water; this remineralization process is considered as an additional part of excretion. Additionally, the conversion of refractory detritus to labile detritus with higher N and P content is associated with an elimination of required nutrients from the water; this assimilation process is considered part of microbial colonization. Note that mass balance might not fully work when concentration of the dissolved nutrients reach zero (Park and Clough 2018).

Toxicants are partitioned between water and the organic compartments. Partitioning to inorganic sediments is not modelled unless the multi-layer sediment model is included. Aqueous exposure of compartments is modelled as sorption / desorption for the detrital compartments, uptake by primary producers, and uptake via gills of animals (sorption and uptake to the whole body is not simulated). AQUATOX models aqueous exposure by 1-compartment non-equilibrium kinetics; however, steady-state partition coefficients and bioconcentration factors are used as constraints to speed up the calculation of internal toxicant concentrations within taxa. The partitioning coefficients are calculated with consideration of physicochemical properties of a toxicant such as K_{OW}, acidity and the degree of ionization. Additionally, animals experience dietary exposure via the assimilation of exposed food. Taxa eliminate toxicants by depuration (excretion) and by conversion of deceived to detrital biomass. Remineralization releases toxicants from detritus to the water column. Toxicants in AQUATOX are subject to a number of different degradation pathways: Microbial degradation of toxicants in detritus and in the water column is modelled similar to detrital degradation and is limited by suboptimal temperature, pH, and dissolved oxygen. Biotransformation to metabolites within taxa is represented by user-supplied first-order rate constants. Biotransformation pathways with multiple metabolites can be modelled that have their own toxicity. Photolysis is modelled based on a near-surface, direct photolysis first-order rate constant and a light screening factor for adjustment in deeper layers. Hydrolysis is modelled based on user-specified neutral, acid-, and base-catalysed reaction rates. Volatilization is modelled using a stagnant two-film model, with the air and water transfer velocities approximated by empirical equations based on reaeration of oxygen (see below). Rates of fate and transport can be modified for the fraction of a basic or acidic toxicant that is considered not dissociated. For effects of toxicants see the following section below.

Oxygen cycling is not explicitly simulated in AQUATOX. The concentration of oxygen in the water column is a function of reaeration, photosynthesis, respiration, decomposition, and nitrification. Dissolved oxygen is by default simulated as a daily average, but the model can also account for diurnal fluctuations due to light-dependent photosynthesis based on hourly simulation time steps. Reaeration depends on temperature, wind, and the amount of existing dissolved oxygen. In streams, reaeration depends on current velocity and water depth. In conditions where ice cover is assumed, as well as in the hypolimnetic segment of the water column in a stratified lake scenario, reaeration is set to zero. In AQUATOX, low concentrations of dissolved oxygen cause lethal and sublethal effects on fish and invertebrates. Sublethal effects decrease the production of gamete biomass and the consumption rates.

pH is modelled dynamically based on the available refractory dissolved organic carbon from detritus, using simplifying assumptions that work within the range of pH = 3.75 to 8.25. pH affects the ionization of ammonia and the hydrolysis and ionization of organic toxicants, which potentially has effects on their fate and toxicity. pH also affects the decay of organic matter and denitrification, and if pH exceeds 7.5, photosynthesis can cause calcite precipitation. Calcite precipitation eliminates dissolved phosphate from the water column that is considered to adsorb to the surface of calcite and buried together with calcite in the sediment.

Sedimentation and resuspension of inorganic sediments can be modelled using one of four available modules that differ in complexity. Suspended sediment in the water column causes lethal and sublethal effects (reduced feeding, dilution of food by sediment particles, stimulation of invertebrate drift, loss of spawning).

Identify the main components and processes in the system

The main components in AQUATOX include the modelled site with its physical characteristics; the abiotic compartments (inorganic sediment, different forms of detritus and the water column) with their contents of suspended nutrients, oxygen and inorganic matter (water column), biomass (detritus) or sedimented inorganic matter (sediment) and with their contents of toxicants and degradation products; and the biotic compartments (taxa) with their contents of biomass, toxicants and degradation products.

AQUATOX includes a high number of processes. Main processes for the growth and degradation of biomass are: loading from inflow and immigration; nutrient assimilation by primary producers; photosynthesis; food consumption and food assimilation by animals; excretion; respiration; death (conversion from organism biomass to detritus, e. g. due to predation, phytoplankton sinking, periphyton sloughing, scour, lethal effects of toxicants); loss of biomass to compartments outside the model domain (insect emergence, fishing, washout, emigration); degradation refractory to labile detritus by microbial colonization; and remineralization of labile detritus to nutrients..

Main processes for the fate of toxicants are: loading; washout; sorption to and desorption from detritus; uptake by primary producers; sorption to gills in animals; assimilation with food in animals; excretion from organisms to detritus and to the water column; release to the water column with remineralization of detritus; and ionization of acidic and basic toxicants because the process rates differ for the ionized and non-ionized forms of a toxicant.

Main processes for the degradation of toxicants are: microbial degradation in detritus and in water; photolysis; hydrolysis; volatilization; biotransformation to metabolites within organisms; and again ionization of acidic and basic toxicants.

How the effects of the chemicals are modelled

AQUATOX relates the direct effects of toxicants to their time-varying internal concentration within taxa (Park and Clough 2014). The ecotoxicology module for the direct effects of toxicants has been designed such that the user needs to provide only the cumulative median lethal concentration for a given toxicant and taxon that has been observed after a given time of constant exposure in a standard acute toxicity test ($LC50_{\text{observation time}}$). The ecotoxicology module then computes the following sequence at the beginning of a simulation (see section 2.11.2.4 for details on the equations used):

- Estimate the internal concentration that causes 50 % mortality when held constant for the observation time in the acute toxicity test (*internal* $LC50_{observation\ time}$). This is done by multiplying the external $LC50_{observation\ time}$ with a bioconcentration factor that is estimated from the physicochemical properties of the toxicant or provided by the user. It is considered that the internal concentration of toxicants with high $\log K_{OW}$ has not yet reached equilibrium by the end of acute tests; therefore, the calculation of the bioconcentration factor is adjusted accordingly to K_{OW} and the observed exposure time.
- Establish an equation for an exponential decrease of the internal LC50 with increasing internal exposure time (based on Mancini 1983). It is thus assumed that the sensitivity of organisms to a given internal concentration (and the resulting mortality) increases asymptotically with exposure time. This equation requires two parameters: First, the rate of decrease in internal LC50 with time (*k2*) is by default assumed to be similar to the elimination rate of the toxicant in a bioconcentration test, since good agreement between both has been observed. This elimination rate is estimated based on the physicochemical properties of the toxicant in the model. Second, the threshold for the internal LC50 after infinite exposure time (*internal* $LC50_{infinite}$) is then estimated by inserting *k2*, the internal $LC50_{observation\ time}$ and the observation time from the acute test in the equation and solving it for $LC50_{infinite}$.

With the established equation, the internal LC50 after an arbitrary period of exposure can be predicted. The modelled increase in sensitivity (≠ effect size!) depends only on the time of previously experienced exposure but not on the experienced concentrations. At each time step of the simulation, AQUATOX then computes the following sequence:

- Calculate the *internal* $LC50_{exposure\ time}$ for the current exposure time, i. e. for the summed-up time steps during which the taxon has been exposed to any concentration of the toxicant in the simulation. The exposure time is limited to the life span in animals, and additionally is reset to zero for all taxa in midwinter when plants are considered to die back and algae are considered to sexually reproduce.
- Specify a Weibull dose-response model for the increase in mortality with the log-transformed currently experienced internal concentration. Insert the $LC50_{exposure\ time}$ calculated above as the inflection point and a default or user-specified shape parameter for the spread and asymmetry of the s-shaped dose-response curve. This equation yields the cumulative fraction of biomass that would have been killed if the taxon was constantly exposed to the current internal concentration for the entire exposure time.
- Calculate the actual fraction of biomass killed at the current time step. First, the existing biomass is differentiated in new (sensitive) biomass that has formed in the current time step and thus never been exposed before, and in old (resistant) biomass that has survived previous exposure. For sensitive biomass, the actual fraction killed is similar to the cumulative fraction killed. For resistant biomass, the actual fraction killed is the increment in the current cumulative fraction killed as compared to the cumulative fraction killed in the previous time step (i.e., resistant biomass is subjected only to mortality that is in excess of the previously experience mortality).
- Subtract the respective fraction killed from the new and old biomass.

Taken together, the overall sensitivity of organisms increases asymptotically with time of exposure (to any concentration) in the model, and effects at the current time step are related to the current sensitivity and the current internal exposure. Over an extended time of constant internal exposure, the toxicant-induced daily mortality will exponentially decrease when the internal LC50 approaches *LC50infinite*. However, daily mortality will never reach zero, because the fraction of fresh biomass formed each time step from growing and reproduction is subjected to the full mortality like at the beginning of exposure.

The magnitude of the modelled direct toxicant effects on taxa (at organism-level) are not affected by environmental conditions or additional toxicants. Therefore, effects of multiple toxicants and environmental stressors are modelled assuming effect addition.

In addition to the decrease in biomass from acute mortality, AQUATOX can simulate the following sublethal effects: reduced photosynthesis, accelerated sinking of phytoplankton, reduced growth, reduced reproduction, increased sloughing of periphyton and increased drift of invertebrates. The strength of sublethal effects is scaled from the calculated acute mortality, using a sublethal:lethal ratio (called *application factor*) that relates the external EC50 for sublethal effects observed in a chronic test to the LC50 for lethal effects in the acute test. Each time step, the sublethal effect size is then calculated from a Weibull dose-response model similar to the cumulative mortality, but the *internal* $LC50_{exposure\ time}$ multiplied with the *application factor* is used as inflection point. AQUATOX uses three separate application factors for the effects on photosynthesis, growth and reproduction. No default values are included, hence the user must provide EC50 values for a toxicant in at least one taxon in order to model the respective sublethal effects. Then AQUATOX calculates the application factors and can use them also for all other taxa, assuming that the sublethal:lethal ratio is constant across species.

Reduced reproduction is modelled as increased mortality of gametes; the fraction of the overall biomass of a taxon which is considered gametes is modelled to be more sensitive than older biomass (due

to the application factor). In animals, reduction in growth is modelled as a combined effect on the consumption rate (20 % of the overall effect size) and on the assimilation efficiency (i. e., an increase in defecation of consumed food by 80 % of the effect size). The effect size calculated for reduced photosynthesis is used also to model an increased sinking rate of phytoplankton. Effects on sloughing of periphyton and on drift of invertebrates lead to an increased washout of both organism groups and are modelled based on the external concentration of the dissolved toxicant in water, because the response is considered to occur immediately. To simulate these effects, the user needs to specify a separate EC50 for drift and sloughing, respectively, as well as a threshold for the maximum loss due to toxicant-induced sloughing and a minimum concentration that affects drift.

How the components and processes are linked

In AQUATOX, each component is affected by multiple processes which are themselves affected by multiple other abiotic and biotic processes and conditions (see description of the conceptual mode above).

Main driving factors of those processes related to biomass growth and degradation are: nutrient or food availability; temperature; pH; dissolved oxygen; toxicant effects (incl. those of ammonia and suspended sediment); light; and stream velocity. Main driving factors of toxicant fate and degradation processes are: temperature; pH; dissolved oxygen; light; and stream velocity.

The magnitude of direct toxicant effects on taxa (at organism-level) depends on the time of previous exposure (to any concentration) and on the current internal (by default) concentration; the direct effects are not modified by environmental conditions or the constitution of organisms (see description of chemical effects above).

2.11.2.4 **Formal model**

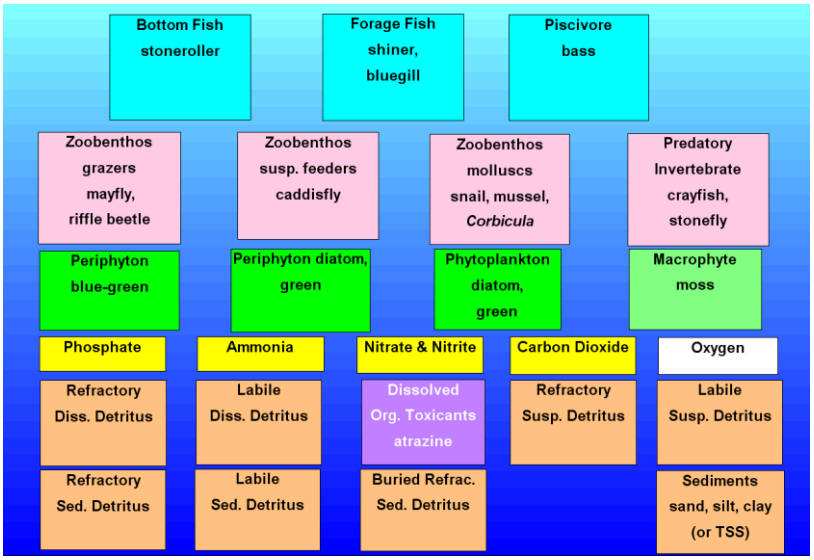
Identification of the model variables

State variables of the basic model include: concentrations of the different nutrients and toxicants, and of dissolved oxygen and inorganic matter in the water column; biomass and amounts of toxicants in each taxon and detrital compartment; amount of sedimented inorganic matter. Optional modules extend the number of driving variables, e. g. to differentiate age or size classes of fish and clam taxa; several ecosystem segments each represented with all or a subset of the above-mentioned state variables; or multiple compartments or organic and inorganic sediments and pore water at the bottom.

Driving variables include: water volume; temperature; light; stream velocity; wind; loadings of water, biomass, suspended sediment, nutrients and toxicants; and water outflow.

Post-processed variables are calculated to ease the interpretation of model results and to assist model validation. They include: various biotic indices (e. g. % EPT) and similarity indices (e. g. Steinhaus index) to compare the community composition in a treatment and a control scenario; various abiotic measures that can be easily measured in the field (e. g. Secchi depth for measurement of turbidity, chlorophyll a content; total organic carbon TOC).

Figure 21: AQUATOX – State Variables in a Typical Model Application



State variables in AQUATOX as implemented for Cahaba River, Alabama (Park and Clough 2018). Note that state variables of toxicants in each biotic and abiotic compartment have not been shown.

Identification of the model parameters

AQUATOX includes more than hundred parameters (typically several hundred), depending on the modelled food web, toxicants and the selection and modules for optional processes (see section 2.11.1.1). The model parameters quantify the ecological processes in AQUATOX that are modelled as transition rates between the state variables. Details can be found in the technical documentation (Park and Clough 2018).

Description of the most important model equations or algorithms

AQUATOX is based on a large set of coupled ordinary differential equations that are numerically solved using a fourth- and fifth-order Runge-Kutta integration routine with adaptive time steps. The reporting time step (that is also the maximum simulation time step) can be set to one day or one hour (when diel fluctuations in light and dissolved oxygen are considered relevant). AQUATOX with all its optional modules comprises more than 450 equations; for details see the technical documentation (Park and Clough 2018). Here we show only the equations to model the direct effects of toxicants in the ecotoxicology module. Equation numbers in Park and Clough (2014) are provided for reference (but some variable names have been edited to improve readability).

Lethal Effects

First, the user-provided *external* $LC50_{observation\ time}$ observed after a fixed exposure time in an acute toxicity test is multiplied with a toxicant-specific bioconcentration factor (*BCF*) to estimate the *internal* $LC50_{observation\ time}$ for the same exposure time.

$$internal\ LC50_{observation\ time} = BCF * external\ LC50_{observation\ time} \qquad \text{Eq. 410}$$

The BCF represents a partitioning coefficient between water and organisms and is calculated from the properties of the toxicant provided in the chemical library. E. g., for algae the *BCF* is calculated as follows:

$$BCF_{algae} = 2.57 * {K_{OW}}^{0.93} * Nondissoc + (1 - Nondissoc) * IonCorr * 0.257 * {K_{OW}}^{0.93} \qquad \text{Eq. 342}$$

With K_{OW} = *n*-octanol/water partitioning coefficient; *Nondissoc* = fraction of un-ionized compound; *Ion-Corr* = correction factor for decreased sorption (0.01 for chemicals that are bases and 0.1 for acids). It is expected that for toxicants with $\log K_{OW} > 5$, no equilibrium between internal and external concentration has been reached by the end of acute tests that typically last for 96 h. Therefore, the calculated BCF is adjusted for the duration of the acute test using an asymptotic function of time and $\log K_{OW}$ (see Fig. 155 in Clough 2014).

In the next step, the following equation is established for the calculation of the internal LC50 after an arbitrary exposure time, assuming an exponential decrease in LC50 with time (Mancini model):

$$internal\ LC50_{exposure\ time} = \frac{internal\ LC50_{infinite}}{1-\exp(-k2*exposure\ time)} \quad \text{Eq. 413}$$

This equation requires an estimation of two parameters, the rate of exponential decrease (*k2*), and the asymptotic threshold of the LC50 after infinite exposure time ($internal\ LC50_{infinite}$). *k2* is assumed to be similar to the elimination rate constant of the toxicant that can be observed in a bioconcentration test, since good agreement between both has been observed (Park and Clough 2014). *k2* is estimated in AQUATOX based on the physicochemical properties of the toxicant. The exact formula varies for the different organism groups; in the following an example for algae is shown:

$$k2_{Algae} = \frac{2.4E+5}{(K_{OW}*LFrac*WetToDry)} \quad \text{Eq. 364}$$

With the following physicochemical properties specified by the user: K_{OW} = octanol-water partitioning coefficient; *LFrac* = fraction lipid (wet weight); *WetToDry* = translation from wet to dry weight. Alternatively, *k2* may be directly specified by the user, who can calculate it offline from the half-life of the toxicant $t_{1/2}$ observed in a bioconcentration test:

$$k2 = \frac{0.693}{t_{1/2}} \quad \text{Eq. 412}$$

The $internal\ LC50_{infinite}$ is then estimated by solving equation 413 for internal $LC50_{inifinite}$ and inserting the internal $LC50_{observation\ time}$ and *observation time* from the acute toxicity test:

$$internal\ LC50_{infinite} = internalLC50_{observation\ time} * (1 - \exp(-k2 * observation\ time)) \quad \text{Eq. 411}$$

With both parameters known, each time step the internal $LC50_{exposure\ time}$ can then be calculated from equation 413 for the current exposure time. The exposure time sums up all time steps at which the taxon was exposed to any internal concentration of the toxicant, but is limited to the life span of the taxon and is reset to zero in midwinter. Then, the cumulative mortality (cumulative fraction of biomass killed) that would be expected after constant exposure to the current internal toxicant concentration for the elapsed exposure time ($CumFracKilled_{exposure\ time}$) is derived from a Weibull dose-response model.

$$CumFracKilled_{expsoure\ time} = 1 - e^{-\left(\frac{internal\ conc.}{internal\ LC50_{exposure\ time}}\right)^{\frac{1}{Shape}}} \quad \text{Eq. 416}$$

With *internal conc.* = the internal toxicant concentration at the current time step; internal $LC50_{exposure\ time}$ = internal LC50 after constant exposure time for the whole exposure time, determines the inflection point of the dose-response curve; *Shape* = shape parameter of the Weibull model, determines the spread and asymmetry of the dose-response curve (*Shape* = 0.33 by default which matched the average of 21 narcotic chemicals in fathead minnows,). If *CumFracKilled* exceeds 95%, then it is set to 100% to avoid complex computations with small numbers. This way, a population actually becomes extinct at high exposure and cannot recover without immigration, which would otherwise be impossible in a model based on differential equations where a tiny fraction would always survive.

The current mortality (fraction of biomass killed at the current time step, *FracKilled*) of the old biomass that has survived previous exposure is then calculated as the increment in the cumulative mortality from the previous to the current time step. Finally, the actual amount of biomass killed at the current time step is calculated from the current mortality for the old (resistant) biomass and from the cumulative mortality for the sensitive biomass that has newly formed in the current time step.

$$Killed = Resistant * FracKilled + Sensitive * CumFracKilled \quad \text{Eq. 417}$$

With *Resistant* = amount of resistant biomass; *FracKilled* = fraction of biomass killed per time step in excess of the fraction at the previous time step (g/g d); *Sensitive* = amount of sensitive biomass.

Sublethal Effects

Sublethal effects are simulated using simple application factors. If the user provides an *external* $LC50_{observation\ time}$ together with an additional acute median effective concentration for a sublethal effect (*external* $EC50_{observation\ time}$), AQUATOX calculates the application factor (*AF*) as the lethal to sublethal ratio:

$$AF = \frac{\text{external EC50}_{observation\ time}}{external\ LC50_{observation\ time}} \quad \text{Eq. 418}$$

Up to three separate application factors (AF_{growth}, AF_{photo}, AF_{repro}) can be calculated per simulated combination of chemical and taxon, using different EC50 values for observed effects on growth, photosynthesis an reproduction ($EC50_{growth}$, $EC50_{photo}$, $EC50_{repro}$). AQUATOX provides no default value for *AF*, i. e. no sublethal effects are simulated unless an *external* $EC50_{observation\ time}$ is provided by the user. However, AQUATOX offers to use an application factor that has been defined for a chemical in one taxon automatically also in other taxa, assuming that the lethal to sublethal ratio is constant across species. This application factor is multiplied with the sensitivity of the taxon to lethal effects at the current time step *(internal* $LC50_{expoure\ time}$) and then inserted into a Weibull module for the strength of the sublethal effect that is similar to the calculation of the cumulative mortality ($CumFracKilled_{expoure\ time}$, see above). For example, the factor for the reduction in growth of animals is calculated as follows:

$$RedGrowth = 1 - e^{-\left(\frac{internal\ conc.}{internal\ LC50_{exposure\ time} * AF_{growth}}\right)^{\frac{1}{Shape}}} \quad \text{Eq. 422}$$

With *Shape* = 0.33 is by default, as in the equation for mortality above.

The overall reduction factor *RedGrowth* is then arbitrarily split between two processes, assuming that 20 % of growth reduction is caused by reduced ingestion of prey, and 80 % is caused by reduced assimilation efficiency (= increased egestion). Reduced ingestion is calculated as follows and enters the equations for the ingestion rate of a given prey by a given predator:

$$ToxIngest = 1 - (0.2 * RedGrowth) \quad \text{Eq. 424}$$

With *ToxIngest* = reduction factor for ingestion due to toxicant effect. Reduced assimilation, expressed as increased egestion, is calculated as follows and enters the equation for defecation of a given predator:

$$ToxEgest = \left(1 - EgestCoeff_{prey,pred}\right) * 0.8 * RedGrowth \quad \text{Eq. 425}$$

With *ToxEgest* = increase factor for egestion due to toxicant effect; $EgestCoeff_{prey,\ pred}$ = fraction of ingested given prey that is egested by a given predator [unitless].

Similarly to the toxicant-induced reduction factor for growth (*RedGrowth*), reduction factors for reproduction in animals (*RedRepro*) and for photosynthesis ($Frac_{photo}$) are calculated from the application factors AF_{photo} and AF_{repro} using Weibull equations. $Frac_{photo}$ enters the equations for photosynthesis and for the accelerated sinking of phytoplankton due to stress. *RedRepro* decreases the survival of gametes (see conceptual model above).

Finally, toxicants in AQUATOX increase the dislodge of periphyton and zoobenthos, and consequently their washout or drift in streams, when respective effect parameters are provided by the user. The effect on dislodge is always modelled based on the external concentration of the dissolved toxicant in water, because the response is considered to occur immediately. Dislodge of periphyton is modelled as follows and enters the equation for periphyton washout:

$$Dislodge_{Tox,Peri} = MaxToxSlough * \frac{Toxicant_{Water}}{Toxicant_{Water} + EC50_{Dislodge.peri}} * Biomass_{Peri} \quad \text{Eq. 427}$$

With $Dislogde_{Tox,Peri}$ = periphyton sloughing due to given toxicant [g/m^3 d]; $MaxToxSlough$ = maximum fraction of periphyton biomass lost by sloughing due to given toxicant [fraction/d; 0.1 by default]; $Toxicant_{Water}$ = concentration of toxicant dissolved in water [µg/L]; $EC50_{Dislogde.peri}$ = external concentration of toxicant at which there is 50% sloughing [µg/L]; $Biomass_{Peri}$ = biomass of given periphyton [g/m3].

Dislodge of zoobenthos is modelled as follows and enters the equation for drift (= zoobenthos washout):

$$Dislodge_{Tox,Zoo} = \frac{Toxicant_{Water} - DriftThreshold}{Toxicant_{Water} - DriftThreshold + EC50_{Dislogde.zoo}} \quad \text{Eq. 427}$$

With $Dislogde_{Tox,Zoo}$ = fraction of biomass subject to drift per day [unitless]; $Toxicant_{Water}$ = concentration of toxicant dissolved in water [µg/L]; $DriftThreshold$ = the concentration of toxicant that initiates drift [µg/L]; $EC50_{Dislodge.zoo}$ = external concentration of toxicant at which half the population is affected [µg/L.

Alternative ecotoxicology module

Alternatively, lethal and sublethal effects in AQUATOX can be calculated also based on the external concentration in water instead of the internal concentration in the biotic compartment. This is suggested for toxicants that are taken up very rapidly or have an external mode of toxicity, such as affecting the gills directly. If this option is chosen, the cumulative mortality after constant exposure is calculated from a 2-parameter Weibull dose-response model

$$CumFracKilled = 1 - \exp(-k * external\ conc.^{Eta}) \quad \text{Eq. 429}$$

with external conc. = the toxicant concentration in the water [µg/L]. The user needs to provide the external LC50 and a slope factor (slope at LC50 multiplied by LC50) from a standard toxicity test. Like with the shape parameter for the calculation of toxic effects based on internal concentrations, the slope factor needs to be entered only once for each chemical is then applied to all taxa. The parameters *k* and *Eta* in the Weibull model are then calculated as

$$k = \frac{-\ln(0.5)}{external\ LC50^{Eta}} \quad \text{Eq. 430}$$

$$Eta = \frac{-2 * external\ LC50 * slope\ factor}{\ln(0.5)} \quad \text{Eq. 431}$$

2.11.2.5 **Computer model**

Description of the model implementation

AQUATOX is the latest in a long series of models, starting with the aquatic ecosystem model CLEAN (see section 0). Release 3 combined all of the various previous versions of AQUATOX for different habitat types into a single consolidated version. This version was written in object-oriented Pascal with the Borland Delphi 2007 development platform and is available for download as a stand-alone program for windows at the USEPA website (https://www.epa.gov/exposure-assessment-models/aquatox). The latest release 3.2 runs on MS Windows XP, 7, 8, or 10.

Checking the computer model for errors, bugs and inconsistencies in the code

The source code is open source (Common Public Licence 1.0). Release 3.0 was subjected to external peer review by Marty Matlock (University of Arkansas), Damian Preziosi (Integral Consulting, Inc.) and Frieda Taub (University of Washington).

Demonstrate that the computer model performs as indicated by the conceptual and formal models

AQUATOX has a long history of applications. According to the USEPA, the model has been reviewed favourably in at least 12 articles and books, and at least 22 peer-reviewed papers have been published (https://www.epa.gov/exposure-assessment-models/selected-publications-aquatox).

2.11.2.6 The environmental scenario

Description of the environmental scenarios, i.e. the environmental context in which the model is run

Setup of the food web and the parameterization in AQUATOX is fully customizable and allows the user to simulate a large variety of freshwater scenarios. However, because AQUATOX uses hundreds of parameters, a model application is typically not started with a new scenario from scratch. Instead, AQUATOX 3.1 Plus comes with 30 built-in scenarios from previous applications, from which the user can select an appropriate example that is then modified with own data according to the local conditions (see section 2.11.1.1). These scenarios cover representative reference water bodies for various freshwater types ranging from artificial tanks to ponds, lakes, reservoirs, streams, rivers and estuaries. The reference water bodies are mainly located in the US, but include also two Danish studies, and comprise water bodies surrounded by varying degrees of agricultural landscape. Each scenario provides a specific food web with typically 1 – 3 surrogate species in each of the modelled trophic levels and organism groups. These species usually include common surrogate species used in the risk assessment of pesticides, such as *Daphnia*, *Chironomus*, a green alga, and fish. Loadings and initial concentrations of pesticides have to be specified by the user. For this task, AQUATOX can be coupled to the HSPF watershed model, but exposure data from FOCUS models could be used as well.

Include description and justification of combination of abiotic, biotic and agro-environmental parameters

This item is specific for a given scenario and cannot be addressed for the model in general. The built-in scenarios are based on applications of AQUATOX to real water bodies that often have been studied intensely for years. Parameterisation typically includes both generic parameterisation using literature data, and site-specific parameterisation using field observations and some extent of calibration. Details on the local biotic and abiotic conditions can be found in the published applications of AQUATOX to the reference water bodies.

2.11.2.7 Parameter estimation

Description of the model parameter estimation

When developing a new scenario from scratch, the user must select one of the following types of ecosystems, which activates type-specific modules: pond, lake, stream, reservoir, enclosure, or estuary. Afterwards, the user must provide the following application-specific settings: simulated timeframe (dates); initial water biochemistry (content of ammonium, nitrate, phosphate, CO_2, oxygen); the initial amount of labile and refractory detritus in sediment and water; specification of the food web and initial biomass; initial concentrations of chemicals to be modelled in each compartment; loadings of water, organisms and toxicants (from inflow, precipitation, fish stockings, point-sources and non-point-

sources); and site characteristics. The required site characteristics depend on the selected type of ecosystem. E. g., for a stream the following information must be provided: length of simulated stream section; surface area; mean depth; maximum depth; mean evaporation; latitude; channel slope; Manning's coefficient (for stream velocity, can be estimated based on stream type); percentage of riffles, pools and runs; constant or time-variable water volume (can be estimated based on Manning's equation), temperature, wind, light and pH; and the initial concentration of inorganic solids (if their simulation is desired). Finally, parameters for the many process rates in AQUATOX and their dependencies on the modelled environmental conditions must be specified.

Typically, not all of the required information is available for parameterization. Therefore, parameterization typically starts with selecting one of the existing built-in scenarios that is then modified according to the current needs. AQUATOX offers five built-in libraries with standard parameter values when creating a new scenario or modifying an existing one. These libraries provide constants and / or driving variables for sites (e. g. depth, daily duration and intensity of light), and parameters for plants (growth and physiology), animals (feeding preferences, growth and physiology), remineralization (e. g. degradation and denitrification rates and their dependencies on environmental conditions), and chemicals (fate and effects).

The libraries for primary producers, animals and for chemicals contain 69, 137 and 69 entries, respectively, but some taxa and chemicals (e. g. chlorpyrifos) have more than one entry that result from different studies on different ecosystem types. The chemical library contains mainly insecticides, herbicides and PCBs. All libraries are thoroughly referenced. The user can modify each parameter value and can specify additional taxa and chemicals that may then be added to scenarios. To create new taxa, information is required regarding the life cycle (e. g. life span, mortality and biomass of gametes); feeding behaviour (e. g. food preferences); physiology (e. g. respiration, excretion, lipid body fraction); responses to temperature, suspended sediments and salinity; sensitivity to chemicals; and others. For new chemicals, information is required on physicochemical properties (e. g. K_{OW}, acidity, degree of ionization); degradation rates in dependence of abiotic conditions; and ecotoxicology (LC50 and EC50 values for the simulated taxa).

A particular difficulty is the parameterization of toxicant effects, as data are typically not available for all taxa in the modelled food web. Therefore, the user can estimate LC50s for each biotic compartment from the Interspecies Correlation Estimation database WebICE (Raimondo et al. 2010). This web application can be linked to AQUATOX and interpolates LC50 values for specific species and toxicants from known LC50 values for related species or toxicants, based on at least 2,081 available significant aquatic interspecies correlations. The uncertainty associated with this estimation can be numerically quantified. The obtained goodness of fit can be utilized within an iterative AQUATOX uncertainty analysis (see section 0) to assess how uncertainty in ecotoxicological data contributes to the overall uncertainty in model predictions.

Parameters estimated from the literature — what are the sources and why are these appropriate?

A detailed bibliography for the data sources of the built-in scenarios with > 100 peer-reviewed scientific publications used for parameterization is provided at https://www.epa.gov/exposure-assessment-models/aquatox-data-sources-documents.

Parameters obtained from calibration — how and why this was done?

AQUATOX provides an automated calibration procedure for those parameters that cannot be well defined due to a lack of data or due to high variability reported in the literature. In the latter case, constrained parameterisation can be used for fine-tuning, as done for a few parameters in an application example for rivers in Minnesota from the technical documentation (Park and Clough 2018). In con-

trast, a few parameters are hardly accessible to measurements and have been subjected as "free" parameters to broad calibration in some model applications. E. g., for periphyton, light saturation levels, maximum photosynthetic rates, light extinction due to self-shading, and the critical force for sloughing have been calibrated in model applications to woodland streams at Walker Branch, Tennessee (USEPA 2001), and to rivers in Minnesota (Park and Clough 2018); see details on these studies in section 2.11.2.9.

AQUATOX can be calibrated to a specific site, i. e. the parameters are iteratively tweaked in order to match the desired state variables to observations at this site. Alternatively, the model can be calibrated to multiple sites by searching for a single set of parameters that produces the best fit of predicted vs. observed state variables across all sites. The calibrated model is then representative for the range of conditions that have been covered by these sites, and modelling results can be considered more robust. This approach was used in the examples of Walker Branch and Minnesota River, where the periphyton-specific parameters have been calibrated by matching predicted with observed chlorophyll *a* values. Aquatox provides a comprehensive statistical analysis for the obtained fit based on the relative bias and ratio of variances in the modelled and observed data distributions (Park and Clough 2018). In contrast to the parsimonious use of calibration in the examples above, Zhang et al. (2013) provides an examples of extensive site-specific calibration to Lake Baiyangdian in Northern China, in which the predicted temporal variation in biomass of different algal groups could be matched to observations very closely.

2.11.2.8 **Sensitivity and uncertainty analysis**

Summary of the sensitivity analysis and identification of parameters with a relatively large effect on model output

In the sensitivity mode of AQUATOX, the user can perform an automated local (= nominal-range or "one-at-a-time") sensitivity analysis for almost all model parameters. Alternatively, single parameters or combinations of parameters can be modified manually for more specific analyses including parameter interactions. A sensitivity statistic may be calculated such that when a given percent change in an input parameter results in the same percent change in the studied, the sensitivity is calculated as 100% (Clough 2014). Typically, the average endpoint values over the entire simulation period are compared. Additionally, a statistical sensitivity analysis (see example below) can be performed using the uncertainty mode of AQUATOX but varying only one parameter at a time.

USEPA performed a comprehensive sensitivity analysis of AQUATOX Release 3.0 with six of the built-in default environmental scenarios to assess the structural integrity of the model and to provide guidance for the analysis of custom scenarios (USEPA 2013). The analysis assessed the sensitivity of various output variables to variation in parameters for process rates (e. g. temperature-dependency of respiration) and for environmental conditions (e. g. water temperature). Additionally, the sensitivity of model output to variation in driving variables for loadings (e. g. daily inflow of biomass and toxicants) and in the initial conditions of state variables (e. g. biomass at the start of the simulation) was assessed. The studies focused on various output variables and covered various water bodies that can be modelled: Algal biomass and abiotic state variables in the hypolimnion of a lake; biomass at different trophic levels in a river without toxicants; biomass at different trophic levels in a pond and a small stream exposed to the insecticide chlorpyrifos; and PCB bioaccumulation at different trophic levels in an estuary. In each study, first, a local sensitivity was performed on various endpoints by changing 45 – 890 relevant parameters (or driving variables and initial conditions). In some scenarios, both a near-range analysis and a far-range analysis was performed, varying one parameter at a time by 15% or 33 %, respectively. Second, particularly sensitive parameters were analysed more closely using statistical sensitivity analysis. In this case, the parameter was varied based on a user-defined probability distribution obtained from literature values.

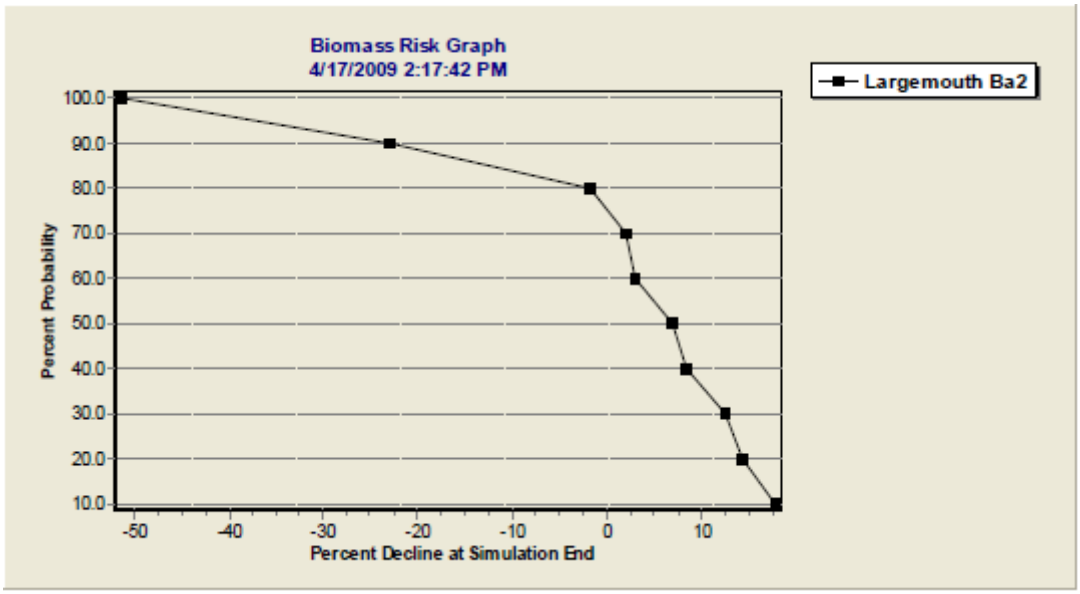
In general, biotic state variables were particularly sensitive to parameters that describe the temperature-dependency of processes and to the water temperature itself, but also to consumption and respiration rates. The authors concluded that water temperature should be parameterized site-specifically, and that care on temperature-dependencies should be taken when calibrating biotic state variables. In addition, phytoplankton was sensitive to the maximum photosynthesis rate and periphyton was sensitive to sloughing. Toxicant fate and effects were highly sensitive to log K_{OW}. Simpler food webs were more sensitive to toxicant-induced food web effects than more complex food webs. Taxa being subject to rapid growth and dieback (e. g. cryptomonads, periphyton, invertebrates) were generally more sensitive to changes than slowly developing taxa (e. g. moss, fish). Parameters that are not of a logarithmic or exponential nature showed an essentially a linear response when extrapolating a 15% change out to a 33% change, indicating that most results of the sensitivity analysis may be extrapolated to a wider range than tested. Feeding preferences were not varied in this analysis, but according to the authors the model was sensitive to the food web setup in previous tests.

Further AQUATOX applications in the scientific literature confirmed that the model is particularly sensitive to temperature and to biological process rates. E. g., Sourisseau et al. (2008) described an independent application of AQUATOX to artificial streams and confirmed the high sensitivity of biomass predictions to changes in parameters related to temperature, maximum rate of photosynthesis and consumption. Similarly, in an application of Zhang et al. (2013) to Lake Baiyangdian in North China, the model was highly sensitive to temperature limitations and respiration rates.

Summary of the uncertainty analysis describing and evaluating the different factors that make the model result uncertain

AQUATOX offers a built-in automated uncertainty analysis (Park and Clough 2018). In the uncertainty mode, the model runs a set of treatment and control simulations with user-defined parameter values and loadings being randomly drawn from Latin Hypercube sampling. Distributions for constant loadings are sampled daily, providing day-to-day variation within the limits of the distribution. Distributions for dynamic loadings may employ multiplicative factors that are sampled once per simulation. For each parameter, a separate sampling distribution can be specified from the triangular, uniform, normal, or lognormal distribution. Sampling distributions for several parameters can be correlated with each other. AQUATOX then prints plots showing the results from a deterministic run, together with the results from the probabilistic runs (e. g. the mean, minimum and maximum biomass of a taxon vs. time). Additionally, AQUATOX offers the plotting of biomass risk graphs, which show the probability that the toxicant-induced decline in biomass by the end of the simulation will exceed a given threshold (Fig. 22).

Figure 22: AQUATOX – Uncertainty Analysis



Biomass risk graph showing the risk from dieldrin to bass in Coralville Reservoir, Iowa, as predicted from a set of probabilistic simulations in the uncertainty mode of AQUATOX. In each of the 10 replicate model runs (depicted as points), the change in bass biomass from the first to the last time step of the simulation has been calculated. The plot shows the cumulative distribution of these declines in all simulations. In one model run, biomass actually increased by ca. 50 %, but in three model runs the biomass decreased by more than 10 % percent., i. e. the predicted risk of >= 10 % decline is ca. 35 %. Note that such a biomass risk graph for the treatment scenario should be evaluated against a second risk graph for the control scenario. Graph reproduced from Park and Clough (2018).

2.11.2.9 **Comparison with measurements**

Description of comparisons of model output with independent data

USEPA presents four model validation reports which focused on different aspects and potential uses of AQUATOX (https://www.epa.gov/ceam/aquatox-model-validation-reports, 31.12.2020). The reported studies used AQUATOX Release 1 so that results for the current model version might deviate to a certain extent.

The first study assessed the prediction of algal blooms and oxygen levels in the heavily eutrophic Lake Onondaga (New York, USEPA 2000). In a first step, AQUATOX was run with an almost default lake scenario to assess the model performance when applied in a screening step analysis. Only the initial conditions, wind, solar radiation, temperature and loading with water and nutrients were adjusted site-specifically using data from Effler (1996). The available annual nutrient loads in 1989 – 1990 were converted to mean monthly loadings estimated based on monthly water inflow from tributaries observed over 28 years. Predicted chlorophyll concentrations and epilimnetic and hypolimnetic oxygen levels were compared to observations in 1989 – 1990. Modelled dissolved oxygen matched observations in the hypolimnion, but missed two epilimnetic oxygen sags observed in summer and autumn. The observed oxygen sag in summer could be related to crashes of algal blooms, as three chlorophyll peaks were observed in summer, while the modelled chlorophyll values remained rather constant during summer. The oxygen sag in autumn was caused by ascending water from the hypolimnion during turnover. In a second step, nutrient loading was refined using daily inflow values observed in a subset of tributaries. This way, peak events of nutrient availability were simulated but did not improve the prediction of chlorophyll and oxygen levels. In a third step, the model was calibrated to site-specific

conditions: High salinity increases stratification in Lake Onondaga; therefore, the depth of the well-mixed epilimnion was reduced to observed values, which increased the effect of ascending hypolimnetic water in autumn. In the simulated food web, blue-green algae were replaced by cryptomonads that form spring blooms in Lake Onondaga, and predatory zooplankton was replaced by rotifers that were set to feed on cryptomonads. The modelled oxygen levels were sensitive to zoobenthos (*Tubifex tubifex*) feeding, which converts refractory detritus formed by crashes of algal blooms into labile detritus that fuels oxygen consumption. Therefore, zoobenthos feeding rates were increased and the initial density of catfish (predator of *Tubifex*) was decreased to observed levels in order to model sediment oxygen demand. These modifications led to quite reasonable matching of predicted with observed chlorophyll and oxygen levels.

The second study assessed the prediction of algal blooms in the large but shallow Coralville Reservoir in 1974 – 1978 that has been built to control floods of the Ohio River (USEPA 2000). The default reservoir scenario was used with only the following site-specific settings: Observed nutrient concentrations in the upstream river were available approx. twice per month, and water inflow and outflow data were available on a daily basis. From these data, daily nutrient loadings were calculated. Further site-specific model setup included weekly temperature values, an average value for wind and solar radiation, and initial biomass of the simulated taxa in 1974. In a first setting, the model was run using a constant water volume by setting the outflow equal to the inflow. As a result, AQUATOX correctly predicted the usually observed range of total algal biomass, but missed the timing when peaks in algal biomass occurred. Additionally, a large algal bloom observed in summer of 1977 was not predicted. In a second setting, daily changes in water volume were estimated from data on pool elevations. This improved the predicted timing of peaks in algal biomass; also the large algal bloom was predicted, though underestimated by 50 %. In concurrence with observations, diatoms dominated over green and blue-green algae in the simulations. The observed and predicted mean biomass of the dominant buffalofish differed by only 12 %; no observations on changes in fish biomass over time were available for comparison.

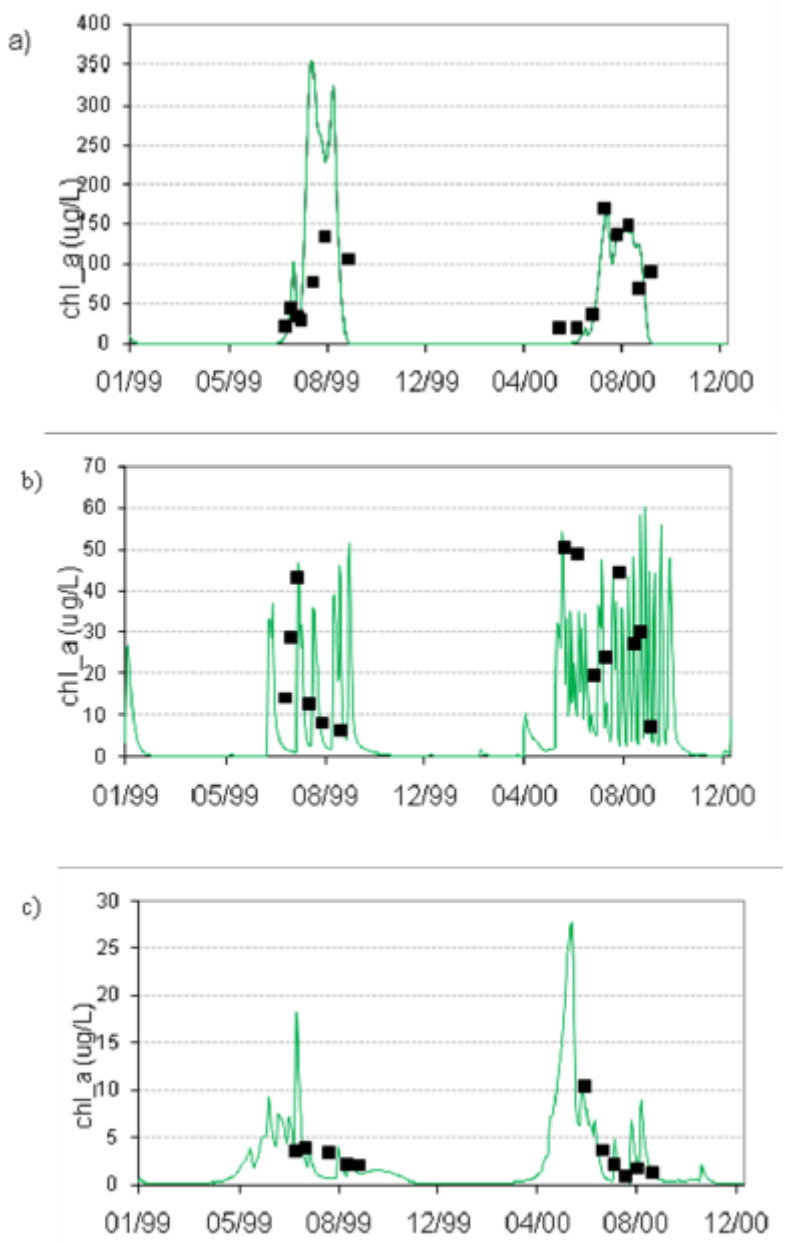
The third study assessed the prediction of bioaccumulation of 16 PCB congeners in the Lake Ontario food web (USEPA 2000). Observed PCB concentrations in water, sediment, plankton, mysids, benthos, and in different fish species (each collected at different dates in the period of 1981 – 1986) were compared with model predictions. Some code and several parameter values of AQUATOX were modified in order to compare the modelling results with simpler models in Burkhard (1998) which assumed constant PCB concentrations in water and constant lipid fractions in the organisms over time. The modelled food web included diatoms, green algae, other algae, detritivorous invertebrates (amphipods), herbivorous invertebrates (cladocerans), predatory invertebrates (mysids), benthic fish (sculpin), forage fish (alewife), small game fish (smelt) and large game fish (trout). Feeding preferences and a constant lipid content for the various organisms were parameterized using site-specific data (in normal applications the lipid content varies over time in AQUATOX). The simulations started with PCBs present only in refractory detritus and in water at observed concentrations. Then the model was run for seven years; after a simulation period of four years, the predicted bioaccumulation factors (BAFs) and biomass reached stable states for all taxa. While the predicted biomass and BAFs showed a clear seasonal pattern in invertebrates and most fish, no seasonality was observed in BAFs for phytoplankton and trout. Predicted BAFs at the end of the simulation period (March 31) were compared with field observations mostly collected in a similar season. Predicted and observed BAFs generally increased with the K_{OW} of the PCBs. While BAFS were overestimated for amphipods (by a factor of ca. 5), sculpin (factor ca. 10) and alewife (factor ca. 8), and underestimated for phytoplankton (factor ca. 8), they were predicted well for mysids, smelt and trout. When compared to the BAFs predicted by the models of Thomann and Gobas in Burkhard (1998), AQUATOX performed best for phytoplankton and mysids, and equally well as the Gobas model and better than the Thomann model for smelt and trout.

The fourth study assessed predicted effects of nutrients, light and grazing on periphyton in a woodland stream at Walker Branch, Tennessee (USEPA 2001). The modelled food web included diatoms, green algae and gastropods. Predicted effects on periphyton biomass were tested with a factorial experiment in stream enclosures and stream-side channels, covering seven weeks in spring 1989, summer 1989, and spring 1990, respectively. The experiment included 20 treatment types representing combinations of high and low levels of nutrients, light intensity (shading) and grazing pressure (snail removal). Poorly defined model parameters were calibrated by tweaking them one after another until phytoplankton biomass prediction for the full dataset (across all treatment types) was optimized. The calibrated parameters included light saturation levels, maximum photosynthetic rates, and light extinction from self-shading for diatoms and green algae, and the maximum consumption rate, minimum biomass for feeding and carrying capacity for gastropods. After parameterization, the predicted and observed temporal patterns in periphyton biomass, as well as the mean and maxima of biomass over the course of the experiment were compared separately for each treatment type and for different seasons (spring and summer). The maximum periphyton biomass was generally overestimated in non-grazed treatments and underestimated in grazed treatments. Under natural conditions (shading, grazers, low nutrient levels), both the predicted and observed periphyton biomass showed a largely stable state over time at 0.35 mg/cm^3, though minor observed variation in biomass were not captured by the model. When grazers were removed, AQUATOX predicted a slow exponential growth in biomass that reached a peak of 6 mg/cm^3 in late spring 1989, followed by rapid breakdown due to sloughing. Because only five observations were available per season (ca. every two weeks), it was difficult to assess the matching of growth patterns. The general trends seemed to be captured, but the timing of breakdown was missed by at least one week. The additional removal of shading pronounced the predicted exponential growth, which was not supported by the limited observational data. Nutrient enrichment under otherwise control conditions (with grazers and shading) slightly increased the stable-state biomass to ca. 0.4 mg/cm^3, which again matched observations though minor temporal variations (0.2 – 0.5 mg/cm^3) were missed. The additional removal of grazers resulted in the prediction of fast exponential growth with up to two breakdowns in spring 1989 at 6 and 4 mg/cm^3, respectively. Observations supported the predicted pronounced growth in spring 1989. In spring 1990, the predicted peaks occurred in between the observations and thus may have been missed. In summer 1989, only slow exponential growth was predicted, though observations suggested rapid initial growth in early summer. The additional removal of shading strongly increased the predicted growth in summer, leading to two biomass peaks at 6 mg/cm^3. Observations confirmed faster growth, though the mean biomass remained lower than predicted. In addition, the model that had been calibrated with data from the stream-side and enclosure study was validated with independent bimonthly data from periphyton on cobbles in the stream channel. The predicted biomass was always lower than the observed biomass, but differed by less than one std. deviation except for a single point. The observational data support the modelled accumulation of diatoms in late summer and autumn of both years, though the timing may have been missed by up to two months, and the predicted subsequent instant decrease due to sloughing could not be confirmed.

The technical documentation of AQUATOX (Park and Clough 2018) reports two further validation attempts. The first study assessed the prediction of periphyton growth, using three rivers in Minnesota for calibration. Time series of driving variables were available for the Blue Earth River and Crow Wing River, and were estimated from interpolation for the Rum River. The model was generally parameterized using literature data, though a subset of parameters with highly variable experimental data was fine-tuned using calibration to data from all streams at the same time. A few parameters such as the light extinction coefficient and the critical force for sloughing of periphyton were broadly calibrated by matching predicted with observed chlorophyll *a* values. The calibrated model was then applied to three dissimilar sites that cover a broad range of nutrient and turbidity conditions in the Lower Boise River, Idaho. AQUATOX overestimated chlorophyll *a* from periphyton in a low-nutrient, clear-water site but matched observations at the other sites reasonably well. Additionally, the calibrated model

was applied to a site on the Cahaba River, Alabama, with modifications to only two parameters: The Crow Wing and Rum Rivers have cobbles and boulders and are more sensitive to higher current velocities than the bedrock outcrops in the Cahaba River; therefore, the critical force for periphyton scouring and optimal temperature for algae were modified based on professional judgement. The chlorophyll *a* content was predicted well and fish and zoobenthos biomasses were predicted reasonably well, though only very few observational data points were available for comparison.

Figure 23: AQUATOX – Model Validation with Semi-Independent Data



Model validation with data from three independent sites to which the model has been calibrated at the same time. Observed (symbols) and calibrated AQUATOX simulations (lines) of chlorophyll *a* in three Minnesota rivers: a) Blue Earth at mile 54, b) Rum at mile 18, c) Crow Wing at mile 72. Note the order-of-magnitude range in scale among the figures. From Park and Clough (2018).

The second study in Park and Clough (2018) assessed the generality of bioaccumulation modelling with the estuarine ecosystem module. Observed concentrations of total PCBs in the water and bottom

sediments from New Bedford Harbor, Massachusetts, were set as constant values in a simulation otherwise parameterized to Galveston Bay, Texas. The predicted PCB concentrations in the various biotic compartments at the end of the simulation were then compared to the observed means and standard deviations in New Bedford Harbor. Good matching of predicted and observed data illustrates that AQUATOX can be used to predict bioaccumulation in a "canonical" or representative estuarine environment without site-specific parameterization.

Few further validation attempts of AQUATOX can be found in the scientific literature. E. g., Sourisseau et al. (2008) calibrated AQUATOX to a single artificial stream and tested the ability to predict the dynamics of biotic and abiotic compartments in other artificial streams with different starting conditions (but none of them was exposed to contaminants). The modelled scenario was comparably simple, including seven aggregated biotic compartments (algae and invertebrates) and only two abiotic compartments (suspended and sediment detritus). After calibration, the initial conditions of the state variables were changed to observed values in neighbouring artificial streams that received the same inflow from a common reservoir (Gave de Pau River). The periphyton chlorophyll a content and the biomass of zooplankton, grazers and predators were in reasonable accordance with observations (≤ 25 % deviation), whereas the biomass of detritus feeders was constantly underestimated.

Demonstration that the model output provides an adequate match to data patterns

The authors of the USEPA validation reports concluded that AQUATOX was generally able to predict abiotic processes such as changes in water levels, dissolved oxygen and bioaccumulation factors well, whereas higher uncertainty was observed in the prediction of dynamics in biotic endpoints such as biomass or chlorophyll a. However, we found no validation of the ecotoxicology module and of the prediction of toxicant effects in general, such as a comparison of the predicted and the observed differences in population development in a control and exposed scenario.

2.11.2.10 Model use

Explanation of how the model conforms to the requirements set in the problem definition

This item can only be addressed in the context of a specific model application for which requirements have been set in the problem definition.

As an example with high relevance to the risk assessment of pesticides, Zhang et al. (2013) applied AQUATOX to assess the risk of PCB pollution to the freshwater community in the shallow mesotrophic–eutrophic Lake Baiyangdian in Northern China. The modelled food web included four phytoplankton, three periphyton, two macrophyte, two zooplankton, one benthic insect, four benthic invertebrate, and two fish populations, as well as detritus. After extensive calibration with lake-specific data, the model closely reproduced changes in biomass of various biotic compartments in the simulated period from March 2009 to March 2010. To assess the risk from PCB exposure, the calibrated model was run in probabilistic mode with different realistic environmental PCB concentrations, and the risk of >= 10 % reduction in biomass of different taxa from the beginning to the end of the simulation was assessed. The initial concentrations in algae and invertebrates were estimated from bioaccumulation in preliminary model runs. At low PCB concentrations (19.46 and 38.92 ng/L), the risk of >= 10 % decrease in biomass was < 0.6 % for all modelled taxa. Concentrations of 77.84 ng/L and above increased the risk of >= 10 % biomass reduction to > 80 % for rotifers, > 1.1 % for copepods, > 4.8 % for chironomidae, > 1.3 % for mussels, > 77 % for carp and > 30 % for catfish. The high risk for top predators is partly related to the modelled biomagnification of PCBs. Conversely, the risk decreased to 0 % for all producers and for crab, shrimp and Asian mud snails, likely due to toxicant-induced release from predation (indirect effect). The predicted risk of biomass reduction by >= 20 % (considered as minimum detectable effect in the field) was compared to experimentally derived NOEC values and

other relevant criteria for regulatory risk assessment in the US; it was assumed that a 0.5 % risk of biomass reduction by 20% is considered acceptable. The experimental NOEC values ranged from 90 to 1,562.5 ng/L (geometric mean = 826.25 ng/L). AQUATOX predicted non-acceptable effects at concentrations of ca. 1 order of magnitude below the NOEC values from standard tests, which is in line with an assessment factor of 10 for chronic bioassays.

Description how the model works (user manual).

AQUATOX is provided via the USEPA website along with a user manual (Clough 2014), a comprehensive technical documentation (Park and Clough 2014, Park and Clough 2018) and a short summarizing fact sheet (USEPA 2018).

Description of the pesticide parameters values used in the model

Parameterization of pesticides is case-specific for a given model application. See section 2.11.2.7 for a description of how pesticide properties can be parameterized in AQUATOX using the WebICE data base.

Description of the specific assessment including a discussion of the most important results

This item cannot be addressed for the model in general. Overall, the USEPA website lists 68 case studies and 25 reviews dealing with AQUATOX that have been published since the year 2000 (https://www.epa.gov/ceam/selected-publications-aquatox, accessed 31.12. 2020). See also the application example of AQUATOX in the ERA of PCBs provided above.

2.11.2.11 Reality/problem — conclusion

Tie in the results from the modelling with the specific protection goal identified in the problem definition section

This item is specific for a given model application.

Can it be established that it is ”clearly established that no unacceptable impact occurs”?

This item is specific for a given model application.

2.11.3 Model Evaluation

2.11.3.1 Evaluation of the problem definition

The regulatory context in which the model is run

AQUATOX has been developed with the support of the USEPA to provide an integrative fate and effects model for the assessment of aquatic ecosystems. The model is intended mainly for use in retrospective risk assessment of anthropogenic impacts such as toxicant exposure or structural changes in sediment load, as well as in the assessment of mitigation and restoration measures. So far, AQUATOX has not been applied to the prospective risk assessment of toxicants. This is partially due to a lack of agreed standard environmental scenarios and impact indicators for effects on the ecosystem level (Lombardo et al. 2015).

The question that has to be answered with the model

AQUATOX can be applied to understand and, in principle, to predict the fate and effects in aquatic ecosystems. Therefore, the model can be used in risk assessment to address the following questions: risk of bioaccumulation in the food web; risk of toxicant exposure to populations in a realistic environmental context; risk of toxicants to non-sensitive organisms due to indirect effects via the food web; risk of toxicants on community structure in aquatic systems; risk of toxicants on aquatic ecosystem functioning and ecosystem services such as food source for fishing and maintenance of high water quality.

The available knowledge and data relevant to the risk assessment question

AQUATOX is a synthesis of various predecessor models that have been developed over > 30 years by experts in ecology and chemistry from the USEPA and different universities. The references for equations and parameterization comprise several hundred peer-reviewed scientific articles and books. Therefore, AQUATOX is based on a broad compilation of knowledge. However, significant knowledge gaps exist particularly regarding the modelling of food web structure and the effects of toxicants at the organism level. Variability in the structure of real food webs is high and information to identify a realistic worst-case scenario for the prospective regulatory risk assessment of pesticides is limited. The limited ecotoxicological information available is complemented in AQUATOX with numerous pragmatic assumptions to extrapolate observed direct effects to different exposure regimes and endpoints. These assumptions have a potentially high influence and result in unknown levels of uncertainty in model predictions.

The outputs required to answer these questions including performance criteria for the regulatory model

Specific protection goals (SPG) have been defined only at organism and at population level. SPG that characterize and quantify acceptable effects of pesticides on whole aquatic communities and ecosystems have not been defined yet. The output of AQUATOX is suitable to address population-level SPG that address the decrease and recovery of biomass in (EFSA PPR 2013). The output would be suitable also to address various expectable SPG for communities and ecosystems once established; this includes the magnitude and duration of pesticide-induced changes in the community composition and diversity, the extent of bioaccumulation, and changes in the chemical water quality.

The species to be modelled

AQUATOX can simulate more than 40 primary producer and consumer species that cover the most relevant guilds and taxonomic groups. The built-in libraries include phytoplankton, epiphyton and macrophytes, shredders, sediment and suspension feeders, grazers and predators from invertebrates and fish from various types of temperate water bodies. Other vertebrates such as amphibians, waterfowl and mammals, and decomposers are missing. Decomposition is modelled as abstract degradation rates that are not directly affected by toxicants, questioning the applicability of AQUATOX for the assessment of fungicides and bactericides.

Requirements for the environmental scenarios to be used in the risk assessment

Assessing pesticide effects on populations in a realistic environment and on whole communities and ecosystems in a Higher Tier ERA requires realistic worst-case scenarios that are representative for the variety of real ecosystems. The built-in scenarios provide comprehensive information on examples for different ecosystem types. When combined with additional knowledge on the variability among ecosystems of the same type, they may provide a good basis for the development of such standard scenarios.

2.11.3.2 Evaluation of the supporting data

Are the data fit for purpose in view of the problem definition?

Formulation and parameterization of the numerous environmental and ecological processes in AQUATOX is based on information from laboratory, but also from field studies. Development and parameterization of the food webs in the built-in scenarios utilized extensive field- studies, sometimes including gut content analyses to assess the feeding preference of species.

Although the quality of the data appears generally high, they still represent case studies under specific environmental conditions. E. g., the model output is highly sensitive to the modelled temperature-dependency of many processes (USEPA 2013). Temperature-dependency has been typically parameterized based on laboratory studies but can be affected by conditions other than those tested for parameterization. Significant knowledge gaps exist also in the connections of food webs (e. g., feeding preferences, limitation of sub-optimal conditions on feeding rates), where parameters needed to be estimated based on exert judgement or calibration.

Has the quality of the data used been considered and documented?

The documentation of AQUATOX does not provide an assessment of the referenced data quality. However, most references cite peer-reviewed scientific studies or data that have undergone quality management within authorities such as USEPA.

AQUATOX offers the estimation of LC50 values for a given toxicant and species via regressions from different toxicants and species. This method is very crude and does not consider specific mode of actions, while the resulting LC50 has a potentially high impact on the model predictions. Running AQUATOX with numerous of such estimated LC50 values may therefore highly increase the uncertainty in predictions.

Have all available data been used? If not, is there a justification why this information has not been used?

AQUATOX references hundreds of scientific publications for parameterization and the development of equations. Because the model has a development history of > 30 years, the development and parameterization of basic equations reference primarily older studies, whereas the more recent applications

reference also new data used for parameterization. To judge whether all available data have been used for a specific equation or parameter requires specific expert knowledge in the respective field.

2.11.3.3 Evaluation of the conceptual model

Are the specific protection goals sufficiently well addressed by the model?

No specific protection goals have been established for the assessment of pesticide effects on the ecosystem level. In principal, AQUATOX is capable of analysing the fate and the propagation of effects of toxicants and other anthropogenic stressors (such as increased sediment loads) on aquatic ecosystems, which was the motivation for the development of the model.

Are the modelling endpoints relevant to the specific protection goal?

AQUATOX can predict effects of pesticides on the overall biomass of populations. The results allow to calculate endpoints such as the highest concentration that will not decrease any modelled population by more than a threshold that is considered to be relevant (NOEC), and the recovery time from pesticide effects for biomass of populations and for biotic indices of the community composition. These endpoints are directly relevant for SPGs for invertebrates and primary producers that are laid out in (EFSA PPR 2013). Applicability in the European regulatory risk assessment for fish is limited, because the SPG for vertebrates refer to individuals and not biomass, but AQUATOX does not differentiate decrease in population biomass due to loss of weight of organisms and due to loss of organisms.

Is the modelling approach justified?

AQUATOX uses a set of coupled differential equations to simulate the flux of biomass in a food web due to a high number of abiotic and biotic processes. This way, the complexity of simulating a whole ecosystem is reduced to an extent that can be run on a normal PC and analysed. However, simulating most populations as simple, non-structured pools of biomass compromises the ability to consider demographic effects in the model. Demographic effects can be relevant for risk assessment: E. g., after lethal pesticide exposure that eliminates sensitive life stages, a population may experience an increased intraspecific competition due to the synchronization of life stages that can delay population recovery (Liess and Foit 2010a). Additionally, the elimination of sensitive life stages (e. g. young larval instars of insects) may result in a temporary loss of food base for a specialized predator species, even if the decrease in the overall biomass of the prey species is low due to the survival of less tolerant life stages that are not subject to predation. Therefore, AQUATOX appears suitable to assess effects on community structure, but results should be considered with care because effects at population structure are disregarded except for fish and clams. A strong and at the same time weak point of AQUATOX is its ecotoxicology module describing the direct effects of toxicants at organism level. The module was designed to require only the minimum and commonly available ecotoxicological information from standard tests. Therefore, AQUATOX calculates effects based on a dose-response approach that is coupled with a 1-compartment toxicokinetic approach and is extended with certain toxicodynamic considerations (see section 2.11.2.3). The model calculates direct effects based on internal concentrations (by default), and is able to predict direct effects that vary with the exposure profile. However, this simplified TKTD approach uses less data but requires more assumptions than more common TKTD approaches (see next section). Together with the built-in default libraries and scenarios for parameterization and setup, this approach enables users to apply the very complex model to numerous toxicants and research questions with a reasonable work load. On the other hand, this pragmatic approach may open the model to potential misuse: While it is relatively easy for a user to generate predictions with AQUATOX, it should be always considered whether the main assumptions on pesticide effects at the organism level may actually hold in a given model application.

Is the conceptual model logical?

Overall, the concept of the model and the representation of environmental and ecological process appears logical. The modelled food web structure is highly flexible and should be assessed separately for each model application.

Assumptions on the effects of toxicants at the organism level in the ecotoxicology module are generally logical but should be supported with additional evidence when using the model for the ERA of pesticides. The concept of the Mancini model used in the ecotoxicology module can be summarized as follows: The fraction of biomass that survived any length of pesticide exposure is principally considered tolerant and will survive also further exposure to the experienced or lower concentrations. Only fresh biomass formed at a given time step is considered not previously exposed and tolerant yet and thus is subjected to the full toxicant effect. However, the tolerance (LC50) is considered to decrease exponentially with increasing exposure time because the capacity of organisms to deal with adverse effects gets depleted.

The ecotoxicology module of AQUATOX actually follows the same considerations as the IT (individual tolerance) mode of action in GUTS (section 2.2): Under constant exposure, the mortality experienced each time step decreases with increasing exposure time due to an accumulation of tolerant individuals in a population. Applying the initial mortality each time step would therefore seriously overestimate mortality after extended exposure, but assuming mortality not to increase with exposure time at all would underestimate real effects. GUTS solves the issue by relating mortality to the accumulated amount of an abstract "damage" at each time step; the rates of damage built-up and repair are subject to study-specific calibration. AQUATOX also considers damage implicitly but assumes that the rate of damage built-up (i. e., the decrease of LC50) can be established from an experimentally observed bioconcentration factor or estimated from the $\log K_{OW}$ of a toxicant and the lipid content of an organism. In contrast to GUTS, the LC50 in AQUATOX remains decreased even after exposure has stopped, until it is reset in midwinter. Therefore, AQUATOX considers carry-over effects in terms of decreased tolerance to additional exposure that can last for several generations in case of multivoltine species. However, suddenly setting back the LC50 at a given date seems not logical; this could be replaced by introducing a daily rate of recovery that could be related e. g. to the generation time of an organism, assuming that carry-over effects will not extend beyond one generation. Additionally, in AQUATOX the decrease in tolerance over time is not affected by the experienced concentrations (different to GUTS). Therefore, exposure to low concentrations is considered to weaken organisms in the same way as high concentrations. Correcting the rate of decrease in tolerance for the experienced concentration would appear more logical, but should be tested with observational data first.

GUTS is limited to acute effects in non-growing, non-reproducing organisms, because the model does not consider that the internal concentration in organisms can increase due to dietary exposure from feeding and is diluted due to growth of new biomass (section 2.2). In contrast, AQUATOX models dietary exposure and growth of fresh biomass, though it does not differentiate between fresh biomass due to organism growth and due to offspring production. Being conservative, fresh biomass in the model is generally assumed to be as sensitive as biomass that is exposed for the first time. This seems logical for fresh biomass from offspring production if no onset of heritable resistance is considered. In contrast, the approach may overestimate effects on fresh biomass obtained from organism growth which is likely as tolerant as the rest of the growing organism.

The effects of multiple toxicants and environmental stressors are treated independently in AQUATOX, applying the concept of simple effect addition. However, it has been shown that toxicants with a similar mode of action, as well as additional environmental stressors, affect organisms synergistically. This is better described by models for concentration addition of toxicants and for stress addition of toxicants and additional steressors (Belden et al. 2007, Liess et al. 2016a); these models are more conservative than effect addition and could be easily implemented. Similarly, exposure to a given toxicant

decreases only its own LC50 but not those of other toxicants in AQUATOX. It seems logical to expand the applied Mancini model such that also exposure to other toxicants (at least those with a similar mode of action) will affect the LC50 of a given toxicant. However, modelling how toxicants and other environmental stressors (suboptimal environmental conditions) together will affect the sensitivity (LC50) of organisms to each of those stressors would require a detailed energy budget model; this seems not feasible due to the limited data available for the various species that could be modelled in AQUATOX.

Sublethal effects are calculated in the same way as lethal effects, using an EC50 that is obtained from the LC50 by multiplication with a conversion factor. However, while the LC50 typically decreases rather exponentially over time, EC50 values do not (Jager and Ashauer 2018b): The time-response of non-reversible mortality (that accumulates the number of affected individuals) is expected to differ from gradual sub-lethal endpoints (that average the degree of response in every individual). Additionally, sublethal effects in AQUATOX always relate to the currently experienced internal exposure and will not persist beyond exposure (only the sensitivity to additional exposure will remain increased). Therefore, AQUATOX assumes immediate recovery from direct sublethal effects when (internal) exposure has ended; chronic and delayed effects of pulse exposure that are known e. g. from invertebrates (Liess and Schulz 1996, Beketov and Liess 2008b) may be underestimated in AQUATOX.

Are the processes included in the model relevant to the addressed issue?

AQUATOX covers most of the processes that are potentially relevant for the distribution of toxicants as well as for the propagation of their effects in aquatic ecosystems However, direct effects on the remineralization of biomass and on microbial degradation of other toxicants are not represented because microbial organisms are not simulated. Unless AQUATOX is extended to subject the rates of re-mineralisation and toxicant degradation to toxicant-induced limitation, the set of direct effects on organisms that can be imposed and propagate to community and ecosystem effects is incomplete in the model, particularly for fungicides and bactericides. Processes and effects related to the characteristics of individuals (e. g. demographic processes, genotypic and phenotypic variation) cannot be addressed due to the nature of the model, except for the discrimination of age/size classes in fish species.

Are the links between different processes to the variables logical?

A particular strength of AQUATOX are the relevant feedback loops that connect the various processes in a logical way to propagate effects. E. g., toxicant-induced mortality will increase decomposition of deceived biomass and thus the biological oxygen demand. Additionally, increased mortality in predators can induce a trophic cascade that puts pressure on producers through the release of grazers from predation, which may intensify oxygen depletion. However, AQUATOX does not link this increased environmental stress to a potential increase in the sensitivity of organisms to the direct effects of toxicants (see discussion on the ecotoxicology module above). Therefore, some relevant feedback loops that would be hard to parameterize remain incomplete.

Are the temporal and spatial scales relevant in regard to the problem definition?

The default reporting time step of AQUATOX is 1 day and simulations can run over several years which is appropriate to assess long-term effects on the ecosystem. The time steps used for calculation are dynamically set using Runge-Kutta methods to avoid problems when solving stiff equations (e. g. for rapid pesticide degradation). The model is not spatially explicit, but can cover a number of connected segments whose volume and dimensions can be set by the user. This enables the user to address spatial heterogeneity within the modelled ecosystem such as epilimnion and hypolimnion in a lake or upstream and downstream sections of a stream.

2.11.3.4 Evaluation of the formal model

Are the most important model assumptions justified by the modeller?

The technical documentation explains all equations of the model. The underlying assumptions are generally well explained for the main ecological and environmental processes but could be improved for the ecotoxicology module.

Are the most important mathematical equations described?

All equations are described and justified in a comprehensive technical documentation for the latest version of AQUATOX Release 3.2 (Park and Clough 2018).

Is there a description of the variables and parameters including their meaning and unit?

The meaning and units of variables are described in the technical documentation. The USEPA website provides a comprehensive documentation of all parameter values used for applications by the US Army Corps of Engineers (https://www.epa.gov/exposure-assessment-models/aquatox-data-sources-documents). Additionally, variables and parameters including their unit and a short description can be exported from the model libraries as excel files.

Is a justification provided if the complexity of the model is appropriate in view of the problem formulation and the available data?

The technical documentation states that biomass-based ecosystem models such as AQUATOX complement individual-based population models (Park and Clough 2014). While population models can focus on single species to assess age- or size-specific effects under certain conditions in detail, ecosystem models can simulate effects on a population within a realistic biotic and abiotic environment, and also effect propagation to whole communities and the abiotic environment. In contrast to many other freshwater ecosystem models such as CATS that can only model concentrations of toxicants in the water column, AQUATOX models the concentration of toxicants in each compartment. This higher complexity allows to integrate the fate and effects of toxicants and make predictions of effects more realistic.

Are references supporting the equations been provided?

Many equations in the current AQUATOX version were acquired from predecessor models. Equations are referenced in the technical documentation.

2.11.3.5 Evaluation of the computer model

Is there a comprehensive and transparent description of the computer model?

A comprehensive description of the computer model is provided in the technical documentation for AQUATXO Release 3.1 Plus (Park and Clough 2014) and Release 3.2 (Park and Clough 2018).

Is the computer code well readable and is it available?

The help function of the AQUATOX implementation for windows provides a link from where the source code can be downloaded. The source code can be read and compiled using the Borland Delphi platform but requires specific programming skills.

Is it demonstrated that the mathematical model is correctly implemented (model verification)?

AQUATOX Release 3.0 was subjected to external peer review and has been applied in numerous studies. It has been reviewed favourably in at least 12 articles and books according to the USEPA website.

2.11.3.6 Evaluation of the environmental scenario

Is the scenario representative for the risk assessment under consideration?

The built-in environmental scenarios cover different types of natural aquatic ecosystems from ponds to lakes and from streams to estuaries, mainly from North America. The representativeness of these case studies for ecosystems across a given geographic and environmental range needs to be evaluated in comparative studies. The environmental scenarios currently available in AQUATOX are likely not directly applicable for EU risk assessment. However, they represent huge collections of field data that may provide a good starting point for the development of standard environmental scenarios for the EU risk assessment of toxicant effects in aquatic ecosystems.

Has the modeler justified the general biological, abiotic and environmental parameters that constitute the scenario?

In each scenario, environmental parameters were adjusted to the observed conditions. However, information on parameters that describe connections in the food web were often scarce so that many of these values had to be estimated based on expert knowledge. Afterwards, often some parameters had to be calibrated to obtain sensible results over a long simulation time (see examples in sections 2.11.2.9, 2.11.2.10).

Has the modeler ensured that the scenario covers the most relevant exposure pathways for the area under consideration?

This item depends on a specific model application and cannot be addressed for the model in general. Depending on the research question, not all applications of AQUATOX include toxicant exposure.

Is the level of conservatism placed into the scenarios appropriate?

This item depends on a specific model application.

2.11.3.7 Evaluation of the parameter estimation

The model parameter estimation has been adequately documented?

The model parameters are described in a comprehensive technical documentation; the basic parameter values were obtained from referenced literature.

Was the quality of the data supporting parameter estimation (literature or experiment) sufficient?

Parameterization of process rates was generally based on peer-reviewed scientific publications or on studies from authorities such as USEPA with own data quality management. Setup of the model to in model applications were based on site-specific data from field studies on physical properties and on nutrient and toxicant loadings.

Were the estimated parameter values realistic?

AQUATOX comprises about 450 parameters that typically describe very specific chemical and ecological properties and processes. The evaluation of realism in all parameter values would require a group of experts from various fields. However, the model has been successfully applied in a large number of studies with no or only minor adjustments of the basic parameters.

Are the data sources sufficiently documented?

The AQUATOX website of the USEPA provides a comprehensive list of literature (> 100 peer-reviewed publications) for the parameterization of the basic model parameters. Additionally, the built-in libraries for species, toxicants etc. provide references for almost all parameters.

2.11.3.8 Evaluation of the sensitivity and uncertainty analysis

Has the sensitivity analysis been adequately documented?

USEPA performed a structured sensitivity analysis of AQUATOX Release 3 to identify the process parameters, environmental parameters, driving variables and initial conditions to which various model output variables are most sensitive (USEPA 2013). See section 2.11.2.8 for details. A comprehensive documentation with conclusions is available at the USEPA website (USEPA 2013).

Is the sensitivity analysis applicable to the situations identified in the problem formulation?

The structured sensitivity analysis was performed across several environmental scenarios to evaluate the model behaviour in general. The sensitivity analysis is therefore applicable to uses within the domain of applicability of AQUATOX, covering the potential use in Higher Tier regulatory risk assessment of pesticides. However, the sensitivity of predicted effects of toxicants to the various forms of model input has not been assessed; this should be performed once the model has been set up for a specific model application.

Have the results of the sensitivity analysis been presented so that they allow identifying the most sensitive parameters?

The documentation provides a summary describing the most sensitive parameters. This is supported by comprehensive detailed information, including a collection of tornado plots showing the sensitivity of selected model predictions to 15 % or 33 % changes in the most sensitive parameters.

Has the uncertainty analysis been adequately documented?

No structured uncertainty analysis has been published. This item depends on a specific model application.

Is the uncertainty analysis applicable to the situations identified in the problem formulation?

No structured uncertainty analysis has been published. This item depends on a specific model application.

Have the results of the uncertainty analysis been presented so that they allow identifying the most uncertain parameters?

No structured uncertainty analysis has been published. This item depends on a specific model application.

Uncertainty is propagated to the model results?

AQUATOX can be run in a deterministic and in a probabilistic mode (section 2.11.2.8). In the probabilistic mode, the user can set the number of model runs and the probability distributions for almost each parameter and state variable and then perform a Monte Carlo analysis. This way, parametric uncertainty, together with assumed variability in environment conditions such as daily temperature and nutrient loadings, is then propagated to variation (variability and uncertainty) in the model predictions. However, structural uncertainty, e. g. due to uncertainty in the most appropriate formulation of processes or in the design of a food web, cannot be assessed with an automatized procedure implemented in AQUATOX.

Have confidence intervals been estimated and has this information been used in further model use?

When AQUATOX is run in probabilistic mode, graphs that plot dynamics in state variables over time show by default the mean, minimum and maximum values and the standard deviation from all model runs. This information is case-specific and can thus be used only for a given model application.

2.11.3.9 Evaluation of the model by comparison with data from independent measurements

Have the performance criteria for the model been predefined in the problem definition?

No performance criteria have been predefined before model validation studies with AQUATOX.

Are the model outputs that are compared relevant in view of the problem definition?

A number of studies have compared predicted and observed population dynamics of various modelled species for a given scenario. However, we found no comparison of predicted and observed differences in biomass or community composition between a control and a similar exposed scenario. In the ERA of pesticides it is most relevant that a model accurately predicts these effects of pesticides. Therefore, the most important output of AQUATOX as an effect model for risk assessment has not been validated with independent data yet. Alternatively, AQUATOX can be applied as fate model. To support this use of AQUATOX, predicted bioaccumulation factors (BAFs) have been successfully tested with independent data. BAFs are relevant model output for risk assessment, though they more used in the US than in the European regulatory framework.

Have the data with which the model is compared been subjected to quality control and is a description of the data available?

The validation reports presented on the USEPA website refer to the older model version AQUATOX Release 1 (USEPA 2000). Some additional studies (see section 2.11.2.9) are available for later model versions. In these reports, data from peer-reviewed scientific publications or from US authority studies have been used for parameterization and validation. Typically, a comprehensive data set was split to first parameterize AQUATOX to a specific ecosystem and afterwards to test predictions for the unused data.

Is the dataset relevant in view of the problem definition?

Data for model validation have been observed in real aquatic ecosystems mainly from the US ranging from streams to lakes and estuaries. They are relevant for the application of AQUATOX to predict effects in such ecosystems, though it should be identified in additional studies to what extent conclusions from the studies can be extrapolated also to European ecosystems.

Is the fit of model output to the data good enough?

AQUATOX has been subjected to a number of studies that either tested predictions of a fully parameterized model with independent data as recommended in the FSA Sci. Op. on GMP (2014b), or assessed the fit of the model after calibration to multiple independent data sets at the same time. The second case is not a classical validation but is also useful to assess whether the model captures the processes that are required to reproduce observed patterns correctly.

In general, the model was able to predict abiotic processes, such as changes in dissolved oxygen and bioaccumulation factors (BAFs), well. Considerably higher uncertainty was observed for the prediction of biotic processes that drive the dynamics in biomass or chlorophyll a (see section 2.11.2.9). AQUATOX was generally able to reproduce observed patterns in biomass, but accuracy was low without site-specific calibration, likely due to high natural variability across sites.

Has the performance of the model been reported in an objective and reproducible way?

The validation reports present the results in an objective and reproducible way.

2.11.3.10 Evaluation of model use

Is a user manual available?

A comprehensive user manual covering the requested items is available at the USEPA website.

Have all aspects of the modelling cycle been documented?

The technical documentation describes the motivation for the development of AQUATOX and how the model works. AQUATOX is a result of > 30 years development including several predecessor models. A detailed description of all steps in the modelling cycle in one publication would be therefore tedious and not very helpful to a user. Instead, all aspects of the cycle have been addressed in the many publications associated with the development and use of AQUATOX. However, the description of the concept and of the underlying assumptions of the ecotoxicology module should be improved in the technical documentation and in the scientific publications on AQUATOX.

Has a summary sheet been provided by the modeller?

This item refers to a specific model application by an applicant of a dossier. However, various summaries of AQUATOX in general are available at the USEPA website, e. g. in the form of a webinar presentation.

When applicable — is the regulatory assessment described?

This item refers to a specific model application and cannot be addressed for the model in general.

Have appropriate conclusions been derived from the risk assessment?

This item refers to a specific model application and cannot be addressed for the model in general. We found no publicly available report describing an application of AQUATOX in the prospective risk assessment of PPP.

2.11.3.11 Evaluation of the suitability of the model for regulatory purposes

Is there a possibility for dialogue between the modeller and the risk assessor?

The USEPA Center for Exposure Assessment Modeling (CEAM) welcomes feedback from users to continuously refine and improve their software, information distribution, and support. Contact information is offered on the AQUATOX website of the USEPA.

Is a version control system implemented?

A version control is implemented. This evaluation addresses the versions AQUATOX Release 3.1 Plus and Release 3.2.

2.11.3.12 Overall judgement

Overall, is the modelling judged suitable for regulatory purposes? Please provide a justification for this overall assessment.

Due to the integration of a fate and effects part, AQUATOX is probably the most comprehensive, but also most complex model available for the assessment of toxicant effects in whole aquatic ecosystems. The unique connection of fate and effects enables the user to study various questions of high relevance that cannot be addressed by other ecosystem or population models, such as the interplay of bioaccumulation, community response and changes in ecosystem functions. AQUATOX simulates most of the ecological processes relevant to address the propagation of indirect effects in ecosystems, such as bottom-up and top-down regulation, biotransformation and the mutual impact of species growth and abiotic conditions.

Due to its long history of development and the support of the USEPA, AQUATOX is much more sophisticated as compared to newer model approaches in regard to user-friendliness, availability of pre-built scenarios and tools for model analysis, documentation and presentation. Against this background, it may be surprising that the model has apparently never been applied in the governmental risk assessment of pesticides. However, this may be justified considering several issues that hinder a successful application for prospective ERA:

First, the ecotoxicology module of AQUATOX relies on numerous assumptions that are necessary to limit the required input information to an extent that is still manageable; however, these assumptions have not been sufficiently studied to justify confidence in the model predictions (see section 2.11.3.3).

Second, AQUATOX focuses on producers and consumers, whereas effects on the community of decomposers cannot be explicitly modelled. This limits the applicability of the model for fungicides and other pesticides with a high expected toxicity to fungi and/or bacteria.

Third, predictions of AQUATOX have not been sufficiently tested with independent data. While predictions on abiotic processes incl. the bioaccumulation of toxicants matched observations reasonably well in a number of studies, higher uncertainty was identified for the simulation of biotic processes. The prediction of toxicant effects has not yet been validated at all. Given the open assumptions in the eco-

toxicology module, a rigorous testing of predicted and observed differences between comparable control and exposed scenarios is essential to gain sufficient confidence for the application of AQUATOX in the prospective risk assessment of pesticides.

Finally, the application of AQUATOX for prospective ERA is limited by the lack of established standard environmental scenarios. Parameterizing AQUATOX to a completely new scenario is tedious and requires extensive data. Therefore, standard scenarios regarding the food web and abiotic conditions need to be developed, and their representativeness for a given ecosystem type must be evaluated to enable the efficient use of models such as AQUATOX.

2.11.4 Qualitative assessment of uncertainties

2.11.4.1 Potential for underestimation of real risk

- Direct effects in the ecotoxicology module are driven by the current exposure at a given time step. Therefore, chronic and delayed direct effects that follow pulsed (internal) exposure cannot be simulated. The assumption of immediate recovery from direct effects after the end of exposure may particularly underestimate sublethal effects and the recovery time of populations.
- Effects of multiple toxicants and environmental stressors are considered to be additive, although they can act synergistically or antagonistically on an individual. Observed combined effects are most often additive (toxicants with different mode of action) or synergistic (toxicants with similar mode of action, environmental stressors), therefore the potential for underestimation is higher than for overestimation.
- Decomposers are not simulated explicitly; effects from reduced decomposition can be therefore underestimated.

2.11.4.2 Potential for overestimation of real risk

- To avoid unrealistically high effects from prolonged exposure, AQUATOX considers biomass that has survived a given internal concentration to be resistant to this concentration in following time steps. Freshly produced biomass is considered as sensitive as biomass that has been never exposed in the model. This assumption ignores the possibility of genetic resistance that might be transferred to the progeny, and possible overestimates the sensitivity of fresh biomass produced from the growth of existing organisms that may be as tolerant as the rest of an organism.

2.11.4.3 Potential for uncertainty in either direction

- Sublethal effects are simulated in the same way as lethal effects after replacing the LC50 with a lower EC50. However, the exposure time-response relationship of gradual sublethal effects is expected to differ from those of the cumulative mortality. Additionally, by default, the LC50 vs. EC50 ratio is considered fixed for a given sublethal effect across all toxicants, irrespective of their mode of action.
- Uncertainties in the parameterization can result in under- or overestimation of real risks. The propagation of uncertainties from parameterization to the predictions can be addressed by sensitivity and uncertainty analysis. AQUATOX is particularly sensitive to temperature-related life processes, consumption and respiration rates, sloughing, the maximum photosynthesis rate, and $\log K_{OW}$ values.
- AQUATOX offers to estimate LC50 values required for model input from regressions using existing data on related species and toxicants. This approach is associated with substantial uncertainties.

3 Applications of Effect Models for the Risk Assessment of Pesticides

3.1 Introduction

In the third part of this report, we reviewed eight case studies for the application of effect models that have been proposed in dossiers for the registration of pesticides in the European Union. The studies were provided by the German Federal Environmental Agency – UBA; they are confidential, but we obtained consent for use in this report with the owning companies prior to publication. Tab. 16 provides an overview of the case studies evaluated. Most studies were dealing with aquatic organisms and only one with mammals, indicating that ecological effect modelling has been mainly used for the refined risk assessment for freshwater organisms so far.

Table 16: Evaluated Case Studies for the Application of Effect Models in the Risk Assessment of Pesticides

Model name	Active substance	Model species	Study
GUTS	Benzovindiflupyr	Fish (five species)	Ashauer (2012)
DEB for Early Life Stages	beta-Cyfluthrin	Rainbow trout (*Oncorhynchus mykiss*)	Zimmer (2017)
IDamP	Bromoxynil-octanoate	Water flea (*Daphnia magna*)	Preuss (2015)
IDamP	Pirimicarb	Water flea (*Daphnia magna*)	Preuss et al. (2009b)
IBM *Chaoborus* population model	beta-Cyfluthrin	Phantom midge (*Chaoborus crystallinus*)	Strauss and Norman (2017)
MASTEP	Pyridalyl	Water louse (*Asellus aquaticus*)	Van den Brink et al. (2007b), Baveco et al. (2012)
MASTEP	Deltamethrin	Water louse (*Asellus aquaticus*)	Verboom et al. (2005)
eVole	Folpet	Common vole (*Microtus arvalis*)	Bastiansen and Meli (2016)

We structured our evaluations according to the checklist in Appendix B of the EFSA Sci. Op. on GMP (2014b) [25]. This checklist has been developed for a structured and comprehensive evaluation performed by risk assessors and was already used for the evaluation of the general models in part 2 of this report. In part 3, we did not subdivide the main sections according to the specific questions from the checklist (as done in part 2), but considered these questions and the background information provided in the EFSA Sci. Op. on GMP (2014b) for the evaluation of the case studies. However, the scheme was developed for the evaluation of both the general modelling approach (the "conceptual", "formal" and "computer" model in EFSA PPR 2014b) and its application in the context of a specific risk assessment (considered as the "regulatory model" in EFSA PPR 2014b). Therefore, we excluded several sections that address the development, implementation and testing of the model in general which has been done independent from the specific case study. Instead, we added a background section at the beginning of each evaluation to familiarize the reader with the general context of the model application, before the specific problem definition was addressed in the following section.

[25] EFSA PPR (2014): Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA journal 12(3): 3589.

3.2 GUTS – Application to Benzovindiflupyr

Evaluation by Tjalling Jager

3.2.1 Problem Definition

In a modelling study of Ashauer (2012) [26], TKTD modelling with GUTS (Jager et al. 2011, Jager and Ashauer 2018b) was used to support the acute risk assessment for fish for the fungicide benzovindiflupyr. Results of the study have been summarized in the draft assessment report (DAR) for benzovindiflupyr by the Rapporteur Member State France (2014) and have additionally been published in the open literature (Ashauer et al. 2013)[27]. Extrapolations were made from the acute laboratory toxicity test to the expected survival due to various FOCUS-SW exposure profiles. This was only done for the two most sensitive fish species (carp and fathead minnow) out of five tested species, where 'most sensitive' was defined as having the lowest 4-day LC50. Furthermore, the model was used to derive a 'margin of safety' in the form of the factor by which the FOCUS profile would need to be multiplied to arrive at 10% mortality due to the chemical at the end of the profile. The GUTS framework has broad support within ecotoxicology, and was recently judged by EFSA to be "ready for use in aquatic ERA" for PPPs (EFSA PPR 2018). The model analysis here largely follows the procedure proposed by EFSA, with two notable exceptions: the focus on 10% effect (where the EFSA opinion works with 50%), and the validation with the early-life stage (ELS) test under constant exposure (whereas EFSA requires validation with pulsed exposure, although for vertebrates, more freedom will be allowed).

3.2.2 Supporting Data

Standard acute toxicity tests were performed with five fish species. The two species with the lowest resulting 4-d LC50s were used for GUTS analysis (i.e., carp and fathead minnow). It is good to stress that this is no guarantee that these species will also be the most sensitive ones when extrapolating over a FOCUS scenario. Using all species in the extrapolation would have been preferable, and this was done for the publication of the results in the open literature (Ashauer et al. 2013), which demonstrated that these two species were indeed also the most sensitive ones under the FOCUS scenarios. However, only two rather extreme exposure profiles were tested with all species (SI of the paper), and the 'sensitivity' of course only considers effects on survival.

For both data sets used in calibration, there was a slight deviation from the test guidelines in that the fish were of slightly larger size than prescribed. However, these data sets were judged in the DAR to be acceptable for the risk assessment. The data are suitable for GUTS analysis, but the number of individuals per treatment is small (only 7), which leads to considerable uncertainty in the parameter estimates (and hence the model predictions) and makes it difficult to identify problematic deviations from the model fit. A simple recommendation to increase the usefulness of toxicity data for modelling purposes is to increase the number of observation times: more frequent observations will generally improve the identification of model parameters without requiring additional animals.

Data from an ELS study with fathead minnow were used as validation data set. These data represent a different life stage, and a longer exposure duration (4 days exposure of eggs, followed by 28 days exposure of the hatched fry, instead of 4 days with juveniles in the acute test: carp of 4.5 cm and minnows of 3.4 cm). Furthermore, in the ELS, the fish will be feeding (either from the yolk or from external food

[26] Ashauer, R., 2012. Modeling fish survival under dynamic chemical stress. Syngenta Report No SYN545192_10325. Confidential Report.

[27] Ashauer, Roman, Pernille Thorbek, Jacqui S. Warinton, James R. Wheeler, and Steve Maund. 2013. A method to predict and understand fish survival under dynamic chemical stress using standard ecotoxicity data. Environmental Toxicology and Chemistry 32 (4):954-965.

supplied) and thus be growing and developing during the toxicity test. Growth, and changes in shape as the embryo develops, may well affect toxicokinetics, and hence provide additional source of differences with the 4-day toxicity test (these processes are not considered in the standard GUTS models). Therefore, these validation data cannot support the model's ability to extrapolate to effects due to long-term time-variable exposure profiles. For that purpose, it is best to have data on the same life stage under the same circumstances (but different exposure pattern). Furthermore, this data set is limited as there is only one treatment with a clear effect on survival. However, these data *can* provide insight whether the calibrated model is likely to be protective for early life stages, and prolonged exposure, as well.

3.2.3 The environmental scenario

The GUTS model does not have an environmental scenario. Survival data from laboratory studies were used to calibrate the model, and the calibrated model was used without further modifications to predict survival from long-term exposure to a FOCUS profile. Environmental temperature may affect the model parameters, as well as food availability (which determines growth, development and reproduction by the animals) and the presence of other stressors. Since these factors are not included in the GUTS model, they will lead to uncertainty in extrapolation from lab to field, but it is good to realise that routine use of LC50s in ERA also ignores these factors.

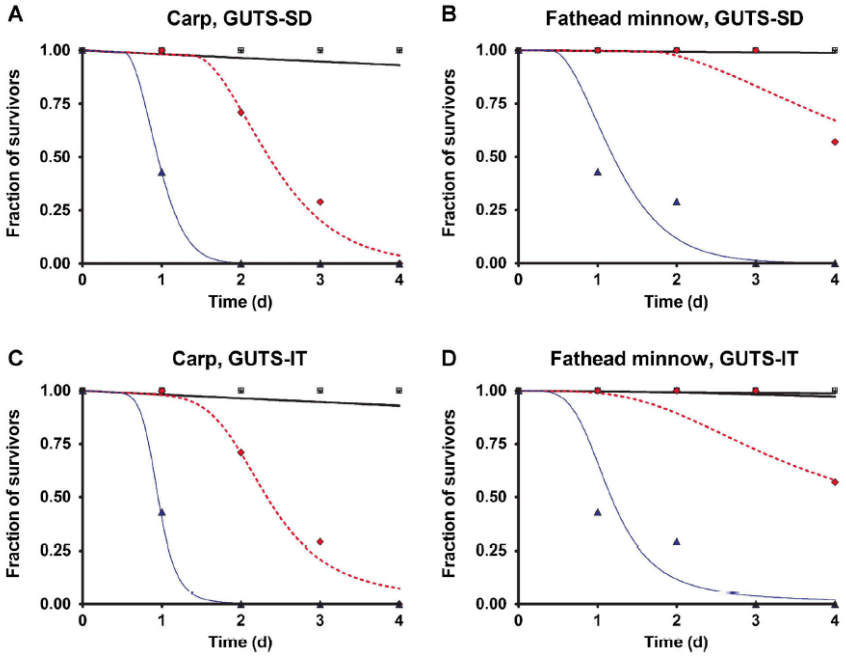
3.2.4 Parameter estimation

The GUTS special cases of SD and IT were calibrated to the raw data from the two selected acute toxicity tests. The implementation in ModelMaker by Roman Ashauer was used for this purpose. The model was calibrated by maximizing the multinomial likelihood, which is the correct method to use for survival data over time (Jager et al. 2011). As optimization method, downhill simplex was used, which is a robust method for rough parameter landscapes as regularly occur for GUTS analyses. Confidence intervals on model parameters were calculated by profiling the likelihood, which is again a robust method for rough landscapes. However, the profiling was performed 'for each parameter separately' (without refitting the others), which implies that the resulting CIs should not be interpreted as 95% CIs (their coverage will generally be considerably less; the correct 95% CIs will be wider, especially when there is correlation between the parameters).

Control mortality was not fitted together with the other parameters. For carp, it was fixed using only the data for the control, while for the other species, a low default was set as no mortality was observed in the tests. This is a valid approach, especially given the low observed mortality in the control as well as in the lowest treatments. Fitting control mortality along with the other model parameters is statistically sounder, but has the downside that, in some cases, an unrealistic background mortality is fitted to compensate for a poor fit in the treatments. According to the information in the DAR, a chi-square criterion was calculated as goodness-of-fit measure, based on the observed and predicted deaths in each observation interval. However, these values could not be found in the results or discussion (and their relevance is limited anyway).

Visually, the fits look quite reasonable, and the two treatments with partial effects provide a reasonable basis for the parameter estimates (despite only few individuals per treatment). Confidence intervals on model parameters were not provided in the DAR, but are included in the publication (Ashauer et al. 2013). I redid the calibrations for carp and fathead minnow and obtained very similar best-fitting parameter sets, but considerably wider CIs. This relates to the limited profiling procedure applied in this analysis, which yields intervals that will have less than 95% coverage. When done properly, the parameter estimates are still reasonably constrained by the data, but the uncertainty is larger, reflecting the small number of individuals in the data set (and only two treatments that produce effects).

Figure 24: Application of GUTS to Benzovindiflupyr – Model Fits



Model fits for GUTS on the two data sets, using two different special cases from the GUTS framework: individual tolerance (IT) and stochastic death (SD). Note that the control treatment is not shown in these plots (the background hazard rate was fitted to the control data, and kept fixed in the fits to the treatment data). Mortality was observed only at the two highest test concentrations for the carp (5.4 µg/L, red diamonds; 10µg/L, blue triangles) and for the fathead minnow (4.4 µg/L, red diamonds; 9.4 µg/L, blue triangles). Graphs reproduced from Ashauer (2012).

3.2.5 Sensitivity and Uncertainty Analysis

A limited sensitivity analysis was provided in the published paper (Ashauer et al. 2013) of the 'one parameter at a time' type. Each GUTS model parameter was changed over its CI, while keeping the others fixed to the best value, and the response of the model output (survival probability at the end of a FOCUS profile, when applying the factor needed to achieve 10% mortality). While this analysis nicely shows that sensitivities differ between the parameters and also depend on the exposure profile, it is unclear what such an analysis adds in the interpretation of the results; this critique basically holds for every type of sensitivity analysis for models that are completely fitted to data (Jager and Ashauer 2018a). Also, changing only a single value at a time might produce misleading results when parameters are tightly correlated.

Parameter uncertainty was not propagated to the model predictions (multiplication factors). In the DAR, coefficients of variation (CV) are calculated for each exposure profile from the four different multiplication factors (two fish species, and two GUTS model). This reflects some of the structural model uncertainties, and the difference between the two species, but does not capture parameter uncertainty (and expressing them as a CV is not particularly useful). Parameter uncertainty is considerable, given the small number of individuals tested, and it would have been useful to propagate it to an uncertainty in the multiplication factors as well. Such a procedure is also proposed in the EFSA opinion on TKTD modelling (although not done for the examples presented in the opinion).

In general, quantification and propagation of uncertainties currently only plays a minor role in ERA (e.g., the CIs on the LC50 are not affecting the TER); how to use the CIs on the GUTS-derived LP10 or LP50 was also not discussed in the EFSA opinion. In general, the CIs will probably be used as a more-qualitative quality control. However, it could be considered to treat them more quantitatively, e.g., by not using the best-estimate of the LP10 but the lower edge of the *x*% confidence interval. The value of *x* should be considered in relation to the AF that is to be applied, to achieve an acceptable overall level

of conservatism. Such a procedure has the benefit that it automatically penalizes 'poor' data sets (data sets that are not capable of identifying the model parameters), while being more lenient for high-quality data sets. Another option could be to use a different AF based on the width of the CI. When using the CIs in ERA explicitly, it is good to consider that these intervals generally only represent parameter uncertainty. What is not captured is structural uncertainty (the fact that the model is always wrong), potential differences between species, and also not the possibility of different sensitivity in a field situation (e.g., due to the presence of other stresses). These uncertainties will be almost impossible to quantify. The 'true' uncertainty will thus generally be larger than covered by the CIs.

Part of the structural uncertainty is addressed in this case study, as the two extreme versions of the GUTS model were both used on the same data sets (GUTS-RED-SD and GUTS-RED-IT, as also required in the EFSA opinion), and since tests from 5 species were analysed (in the SI of the paper). However, a number of other uncertainties remain, such as the influence of other stressors/environmental factors, growth and development of the fish, and the simplifications in the model (discussed in more detail in the evaluation of the general model, see section 2.2). These uncertainties also hold for the dose-response curves that are currently applied in risk assessment.

A proper quantification and propagation of parameter uncertainty is an important aspect of modelling, and should be the general procedure for the use of TKTD models such as GUTS in ERA. However, it is good to realise that GUTS is not adding uncertainty into the risk assessment but rather makes it visible (at least to some extent). Using an LC50 on a peak concentration from a FOCUS profile has considerably greater uncertainties associated with it, but these are not (and cannot) be readily quantified (the results thereby suggest a much greater degree of accuracy than is warranted). Furthermore, uncertainty in the FOCUS calculations is not quantified at all, and may be considerable. The best way of dealing with uncertainties in ERA requires some thought, and it is good to realise that we are holding the TKTD models to higher standards in this respect than the models currently used in ERA (Jager and Ashauer 2018a).

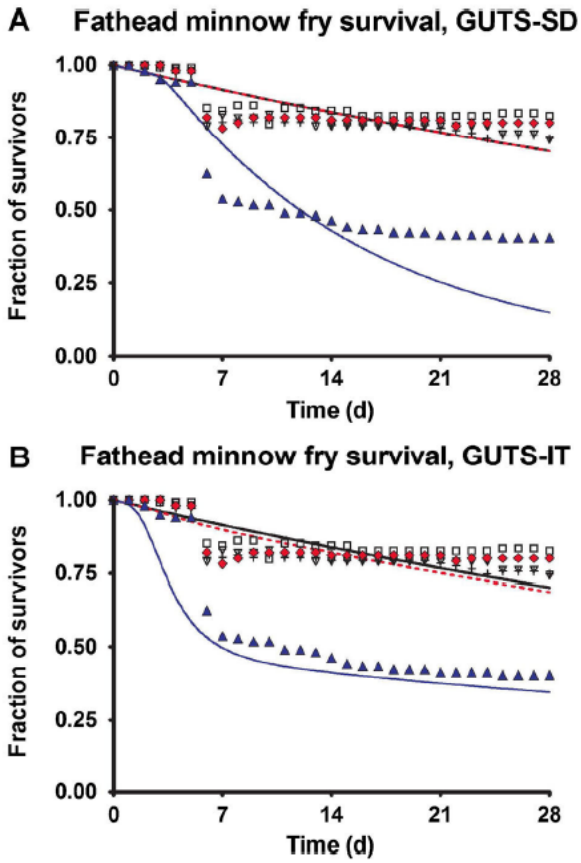
3.2.6 Comparison with Data from Independent Measurements

Data from an ELS study with fathead minnow were used as validation data set. These data were not used for calibration. Background mortality was calibrated using the control data of the ELS; the other parameters were taken without modification from the analysis of the acute test data for fathead minnow. The exposure of the egg stage was ignored in the validation; the comparison starts at hatching, assuming that the hatched fry was clean. The reason is that the egg stage is probably very different from the fry stage in terms of uptake and elimination of compounds. If the chemical is taken up in appreciable amounts in the egg, it is expected that the model would underestimate the observed toxicity. If the model matches the data well, this can be interpreted as confirmation for the assumption that the egg does not accumulate chemicals, or that the model (calibrated on juveniles) is conservative for the early life stages. The assumption of negligible uptake of chemicals in eggs seems to work well in this case, but should be carefully considered in general (for other compounds and other species, this assumption may fail).

A number of factors are different between the calibration data and the validation data: the pre-exposure of the egg stage, the life stage (and thereby, a. o., body size and presence of yolk), the exposure duration, and the introduction of food in the test system when the yolk sac has been absorbed (and hence growth and development of the organisms). All these factors are known to affect toxicokinetics, and might also have affected toxicodynamics. This hampers a straightforward mechanistic interpretation of any observed deviations between the data and the model predictions: poor performance does not mean that the model is inappropriate, and good performance does not mean that the model is appropriate. However, this validation can be seen as a check whether the modelling results would also protect early-life stages (which are often considered to be more sensitive) for longer exposure durations. Interestingly, the nominal concentrations in the ELS were used for the model predictions, while

mean measured were used for the calibration study. This latter observation is unexpected as mean measured concentrations are reported for this test elsewhere in the DAR (they are at maximum 10% less than the nominal concentrations, so no large bias is expected).

Figure 25: Application of GUTS to Benzovindiflupyr – Model Validation



Predictions from the model are compared to data for early-life stages. The drop in observed survival at day 5-6 coincides with the transition from yolk-feeding to external feeding. Note that *t*=0 marks hatching, but that exposure started with the fresh eggs. Blue triangles mark 4 µg/L treatment; other symbols show the treatments with lower concentrations (0.25-2 µg/L; again, the control is not plotted but was used to calibrate background mortality). Graphs reproduced from Ashauer (2012).

The model predictions are quite consistent with the data (and somewhat conservative, although that may disappear when measured concentrations would have been used), which indicates that model simulations are also protective for early life stages under prolonged exposure. Furthermore, the model predicts a clear separation between exposure concentrations that induce additional mortality and those that do not. The data from the ELS support this separation, and thereby also provide some assurance that there will not be additional mechanisms of action that lead to mortality after prolonged low-level exposure. There is a clear drop in survival at 5-6 days post hatching, which coincides with the time at which the fry starts to feed exogenously. This is not captured by the model, but might indicate a change in TK (e.g., importance of feeding, or increased activity, for uptake of the compound).

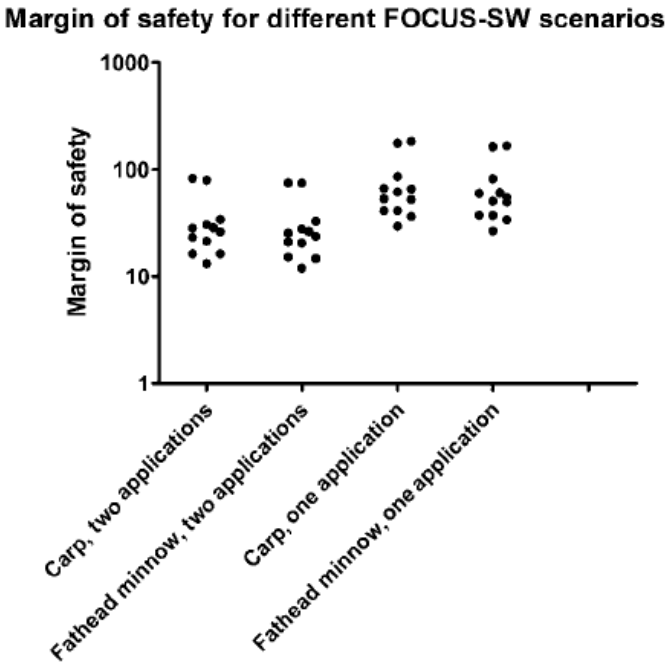
Another major limitation of the validation data set is that it concerns a test with constant exposure, just like the calibration data. The ability of the calibrated GUTS models to predict survival as a result of pulsed exposure (as in the FOCUS profiles, which is the purpose of this analysis) is therefore not demonstrated in this case (even though there is some support for this type of extrapolation from GUTS studies with other compounds and species). The EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) gives specific requirements for validation studies, a. o., the presence of data sets with pulsed exposure

(two scenarios, three treatment levels per scenario), although these requirements are loosened for vertebrates: "an expert evaluation needs to identify the suitability of the validation data set on a case-by-case basis." It is good to stress that this particular model analysis was performed and published well before the EFSA recommendations appeared. As a final limitation, it is good to consider that the validation was done for only one of the two sensitive test species.

3.2.7 Model Use

The calculated safety margins ranged from a factor of 26 to 500. This implies that the FOCUS profiles need to be multiplied with this factor to induce 10% mortality in the calibrated GUTS models. However, no confidence intervals on these factors were calculated. Note that the EFSA opinion focuses on the factor to yield 50% mortality at the end of the profile.

Figure 26: Application of GUTS to Benzovindiflupyr – Margin of Safety for Different Scenarios



Margin of safety for different FOUCS-SW scenarios. Safety values calculated as the factor by which the FOCUS profile must be multiplied to reach 10% mortality at the end of the profile. Note that each clump of 12 points represents 6 FOCUS profiles calculated by two models: SD and IT. Graph reproduced from Ashauer (2012).

Interestingly, the LP10 values in the published paper differ from the ones reported in the DAR: in the paper the safety margins (or LP10) ranged from a factor of 12 to 184, while the DAR provides considerably higher values of 26 to 500. An update of the FOCUS calculations is the most likely explanation (in the DAR, there are also more profiles tested).

In this case study, the RMS concluded that "the use of a 10x safety margin [on the lowest LP10] is acceptable for acute risk to fish, considering that there are data for 5 fish species." The EFSA opinion suggests a safety margin of 100 when using an LP50. The reduction in safety factor from 100 to 10 thus includes considerations for effect level (50% to 10% mortality) and expected differences in sensitivity between fish species. Since the interest of RA is in small effects, it is better to use the model to go from large to small effects (and hence use the LP10), rather than using a safety factor to cover this aspect. The model can calculate both LP10 and LP50, with CIs, and there is no technical reason why one should be preferred. It furthermore would make sense to include the width of the confidence interval into the assessment, e.g., by using a 'lower edge' (or quantile) rather than the best estimate for the

LP10. However, as part of the uncertainties will then be addressed by the model, a lower quantile of the LP10 should be judged against a lower safety margin than the best-estimate of the LP50. In setting an appropriate safety margin, it could also be considered to increase it when validation is limited (as in this case study), and it would be good to consider the expected uncertainty in the exposure estimate (which carries substantial uncertainty that is not quantified). In any case, it would be prudent to consider the safety margin in relation to the factor that is used for an assessment based on a 4-day LC50 and a peak concentration (a 'model' that carries substantially more uncertainty, and is not necessarily worst case).

3.2.8 Overall Judgement

The use of profile-specific multiplication factors is an excellent way to judge the safety margin for each exposure profile. It is also an excellent tool to rank FOCUS profiles (and thereby application types of the pesticide) in terms of risk, when referring to lethal effects on individuals (GUTS does not consider sub-lethal effects that may affect population sustainability at lower exposure levels). This approach was also adopted in the recent EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018). The GUTS analysis in this case study makes optimal use of the available survival data and their time dependency, and of the exposure patterns over time. The data basis for the analysis is limited, which will generally be the case for vertebrates. Given the experience gained with the GUTS model and its predecessors, using the model in this manner is far superior to the use of a 4-day LC50 and mean or peak exposure concentration. However, this type of application would be served by some general case studies with fish to show the level of accuracy with which GUTS is able to extrapolate from acute toxicity tests to longer-term pulse exposure scenarios (currently published validation studies were done with invertebrates).

The calibration can be criticised for the incorrect procedure to derive CIs on model parameters (leading to CIs with less than 95% coverage), and the fact that parameter uncertainties were not propagated from the calibration to the multiplication factors. This would have been helpful, especially in view of the considerable uncertainty in the parameter estimates (a direct consequence of the limited number of fishes in the toxicity tests). The conclusions from the analysis would have been stronger if a validation study with pulsed exposure would have been available (for the same life stage as used for calibration, using measured rather than nominal concentrations), rather than an ELS with constant exposure. Even though this validation provides additional reassurance for coverage of early life stages, and extrapolation to longer exposure durations, it cannot be used to validate the model for extrapolation across exposure scenarios (as discussed above). Nevertheless, this is a useful analysis, and, when appropriate CIs on model predictions would have been included (i.e., propagation of parameter uncertainty), constitutes the best-available technique for using the information in the standard toxicity tests that have been performed for this compound.

Clearly, there are a number of uncertainties, resulting from the model assumptions, when extrapolating from acute toxicity tests to a FOCUS profile under field conditions. The analysis assumes that the model parameters established under constant exposure in the laboratory test are also applicable to (and remain constant for) fish exposed in the field to time-varying concentrations over much longer durations. In the field, fish will grow and ultimately reproduce, which will affect their TK and possibly their TD as well (although this will likely act to decrease toxicity, relative to the toxicity test). In this study, only one life stage (juvenile) was used for calibration, though this was shown to provide a reasonable coverage for early life stages (for one species) as well. Furthermore, no additional mechanism of action should kick in over longer exposure durations or repeated exposure (the results of the ELS provide some reassurance against such additional long-term effects). And, the analysis does not account for multiple stressors, although the fish in the acute toxicity test are not fed (and hence experience food stress as well as toxicant stress). The general uncertainties of GUTS extrapolations are discussed in more detail in the general analysis of GUTS (section 2.2), with a qualitative expectation of whether they will tend to increase or decrease the risk.

The uncertainties that could act to increase risk need to be captured by the safety factor, which needs to be carefully considered (as discussed above). In this case study, a factor of 10 was judged appropriate by the RMS for the LP10 (given this is for the most sensitive from five tested species). Whether that is sufficient or not is impossible to judge for me, and hinges on the question of what one likes to protect and how much certainty is needed, as well as on the uncertainties in the exposure predictions (which are not quantified). The limited validation (no testing with pulsed exposure) could perhaps be weighed into the safety factor, as well as the choice for an LP10 or LP50, and the uncertainty in these estimates.

Finally, it is good to stress that GUTS analyses only concern effects on mortality and hence cannot guard against sub-lethal effects, nor ensure the sustainability of populations. GUTS extracts the absolute maximum amount of information available in acute toxicity tests, but these tests are inherently limited in their information content.

General Recommendation

The EFSA opinion focusses on a specific workflow for GUTS models: calibration using data under constant exposure, and validation using data under pulsed exposure. However, there is no particular reason why the opposite won't be useful. If the model is a good representation of reality, both calibrations will yield very similar parameter estimates, and the model can be fitted on both data sets together to produce a stronger identification of the model parameters (i.e., more certainty in the model predictions). In general, it is best to minimise the distance that the model would need to extrapolate across. For an exposure profile with just a few peaks, calibrating on pulsed exposure data is likely to produce a more useful model than calibrating on data for constant exposure. For a profile with longer sections of more-or-less constant exposure, the reverse would be preferable.

3.3 DEB for Early Life Stages – Application to beta-Cyfluthrin

Evaluation by Tjalling Jager

3.3.1 Background

In the frame of the re-registration of the active substance beta-cyfluthrin, Zimmer (2017)[28] applied a DEB model to investigate how the exposure profile influences effects in the various developmental stages of the rainbow trout. Results of the study were additionally published in the open literature (Zimmer et al. 2018)[29].

Firstly, it is good to stress that the model application in this dossier is not 'DEBtox' in the classical sense, which applies a simplified DEB model (Kooijman and Bedaux 1996a, Billoir et al. 2008b, Jager and Zimmer 2012). However, in the broader sense, it is a 'DEBtox' approach, as a DEB-based model (in this case the full, so-called 'standard', DEB animal model) is applied in ecotoxicology. This potential for confusion was discussed in more detail in the general evaluation for DEBtox (see section 2.3). Furthermore, the model application presented in this dossier (for early-life stages) is a rather atypical one, which does not have much of a track record. Most of the published DEBtox applications have focussed on the analysis of data for growth and reproduction in juveniles/adults (e.g., using data from the 21-day reproduction test with *Daphnia magna*), and most of them have applied the classic (simplified) DEBtox model. The application evaluated here, however, deals with the analysis of results from early-life stage (ELS) testing with fish. These tests start with fertilised eggs and continue until they reach the juvenile (free-feeding) stage in the controls. The initial stages feed on the yolk provided by the mother in the egg. Yolk, and the switch in feeding mode when yolk runs out, is not considered in the classical (simplified) DEBtox model, which therefore cannot deal with the embryo stage. However, the standard DEB model was used here, which does cover the entire life cycle. The recent EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018) presents the classic DEBtox model in Chapter 5, but evaluates this particular early-life stage study with the standard DEB model in Appendix G (according to their proposed checklist for DEBtox models). Therefore, this model application is within the scope of the EFSA opinion, but not the typical application envisaged.

This study is related to some exploratory work that has been done on toxicity for embryos and early life stages for zebrafish (Augustine et al. 2012) and pond snails (Barsi et al. 2014). Even though these two species are very different, DEB models are in principle generic: species should differ in their model parameters for the energy budget, and only to a much lesser extent in model structure. Therefore, findings for snails are relevant for modelling fish as well. The standard DEB model applies to the whole life cycle, and thus also to embryos and early life stages (see e.g., Zonneveld and Kooijman 1993, Augustine et al. 2011). The closely-related (but simpler) DEBkiss model has also been used for analysis of embryonic development (Jager et al. 2013, Barsi et al. 2014, Jager et al. 2018). As noted above, the classic DEBtox models are restricted to the juvenile/adult life stages, and are incapable of dealing with the yolk-feeding life stages (eggs and yolk-sac fry). Therefore, for this purpose, a less-simplified model is required, which in turn increases the information needed about the species. Classic DEBtox models can generally be parameterised using only information from the toxicity tests; for the standard DEB model, as used here, that is impossible. Application thus requires a full set of basic parameters (those

[28] Zimmer, Elke I. 2017. Rainbow Trout: Dynamic Energy Budget (DEB) model simulation to extrapolate effects of beta-cyfluthrin on growth from constant exposure to time-varying exposure. ibacon GmbH, Arheiliger Weg 17, 64380 Rossdorf, Germany. Report No. 111871520. Sponsor: Bayer AG – CropScience Division, Effect Modelling. 40789 Monheim Am Rhein, Germany. Sponsor Ref. No. MOAA0004. Confidential Report.

[29] Zimmer, Elke I., Thomas G. Preuss, Steve Norman, Barbara Minten, and Virginie Ducrot. 2018. Modelling effects of time-variable exposure to the pyrethroid beta-cyfluthrin on rainbow trout early life stages. Environmental Sciences Europe 30 (1):36.

governing the response in absence of toxic stress) from an on-line parameter collection (add-my-pet, see Marques et al. 2018).

The basic principle of 'DEBtox' is that toxicants, once inside the body, will affect one or more of the energy flows in the organism (e.g., maintenance or growth costs). This principle holds for embryos as well as adults. However, for embryos, additional care is needed to properly account for toxicokinetics: this process is likely very different for eggs than for free-swimming stages, and the dynamics of yolk absorption may preclude simple one-compartment kinetics. Furthermore, many relevant embryonic endpoints cannot be (directly) captured by following bioenergetics, such as malformations and endocrine effects. In short, the application of DEB-based models to analyse and predict toxic effects on embryonic and larval life stages is logical but does not yet have the body of background research and useful case studies that the application to the juvenile-adult stages has. This particular case study should thus be treated as a novelty, and the conclusions should not be transferred to application of 'DEBtox' in general.

3.3.2 Problem Definition

The application of the DEB model was only submitted as "additional information" in the dossier, and not used in the actual risk assessment. The aim of the analysis was, as the author states: "… to understand and explain the effects of beta-cyfluthrin on juvenile rainbow trout (*Oncorhynchus mykiss*) using a TK-TD modelling approach." Followed by: "A Toxicokinetic – Toxicodynamic (TK-TD) model can be used to mechanistically explain the differences between the observed effects in the two ELS studies." The aim thus seems to be limited to presenting a TKTD modelling approach for ELS studies, and to demonstrate how this can be used to explain the differences between two specific studies (one with constant exposure and one with pulsed exposure). More specifically, the following questions were posed by the author: "How do relevant life-history parameters of the rainbow trout change with environmental conditions, e.g. different food conditions or temperature? What is the most sensitive life stage to the compound under study? What effect would a different exposure pattern show? E.g. higher or longer peaks." And, at the end of the report: "In the future, the validated model will be used in order to predict the survival and growth of early life-stages of the rainbow trout to beta-cyfluthrin and its formulated products under various realistic peak exposure scenarios (e.g. focus scenarios)." It thus appears that the more long-term aim is to use the model for extrapolations to time-varying exposure, and possibly other environmental conditions.

The species addressed is rainbow trout. The publication (Zimmer et al. 2018) explains that this species was selected because it is a standard test species (ELS studies were available), because of its economic importance, and because it was identified as one of the most sensitive species based on acute data. In principle, the model results could be used to make predictions for other species, by taking the basic DEB parameters for a related fish species and assuming the toxicological parameters are the same. However, the degree of accuracy of such predictions is currently unknown.

3.3.3 Supporting Data

Two types of data are being used in this study: data underlying the add-my-pet entry for the basic parameters (governing the life history in absence of toxic stress) and the data to parameterize the toxic effects in the context of the DEB model. The add-my-pet entry provides references for all of the data used (all publicly available, and mostly from peer-reviewed publications or general databases): https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Oncorhynchus_mykiss/Oncorhynchus_mykiss_res.html. This particular entry includes a substantial amount of relevant experimental data for early-life stages: age at hatch vs. temperature, yolk weight and yolk-free-dry weight vs. time up to 90 days, and several growth curves for juvenile fish.

The toxicity part of the model is parameterized and tested using two ELS studies, which are not publicly available, and not available for this evaluation, but described in detail in the report (and in more

detail in the supporting information from the publication). One experiment is referred to as the calibration data set, which was performed under constant exposure. The mean measured concentrations were used for modelling, and differed considerably from the nominal concentrations. The second experiment is referred to as the validation data, comprising pulsed exposure in a static setup. Two pulse dosing events were given, with 14 days between the events. Three cohorts were used such that the pulses hit a different stage of the life cycle (egg, sac fry, and swim-up fry). Figure S4 in the SI of the paper shows how the exposure concentration decays after the pulse (DT50 around 4 hours). The paper (methods section and graphs) explains that the actual pulse shape (based on the measurements over time) was used as input for the model calculations. The validation test was performed including sediment, which may have led to additional differences in bioavailability compared to the calibration study. Furthermore, the validation data set is for beta-cyfluthrin, but the calibration is for cyfluthrin. According to the publication (Zimmer et al. 2018):"Cyfluthrin is a mixture of four isomers (two cis and two trans) while beta-cyfluthrin only contains the two active isomers (one cis and one trans)." It would appear that some correction of the measured concentrations would be needed to make them comparable between the two studies, but this was not explained.

It is good to stress that the standard test protocol for ELS was never intended for fitting TKTD models, and the data basis from these toxicity tests is extremely poor. There is only body size/mass at the end of the test, although survival is scored more regularly (weekly). All of the information on the dynamics of the effect is thus based on survival; it is assumed that sub-lethal effects adhere to the same dynamics, i.e., the same 'dominant rate constant' governs both lethal and sub-lethal effects. This is a logical assumption, but there is currently insufficient evidence to support this as a general rule. Both tests included observations on timing of events: time to hatch, time to reach the swim-up stage (complete yolk absorption). The timing of these events could have been compared to the model predictions, which may have provided support for the basic parameterization, and for the assumption that there are no effects on the non-feeding stages. The validation data set additionally contained measurements of feeding behaviour after swim-up (although results are not provided). This additional endpoint could perhaps have been compared to the decrease in scaled food level (f) due to the toxicant as predicted by the model. A match between the two would strengthen the selection of the feeding mode of action (MoA) as the dominant one. However, the level of effect in the validation test was too small to be very useful anyway.

The author writes in the modelling report that "Study results suggest that constant exposure to cyfluthrin induced behavioural effects that reduced the ability to feed." However, it is not clear what the basis of this claim is and whether it excludes the possibilities for other MoA's. The paper refers to Groh *et al* to support this statement (however, the wrong paper was cited, it should have been part II, which indeed stresses the link between pyrethroids and reduced food intake in fish).

The author states that all available data has been used, and that the quality of the data has been considered and documented. From this report, the latter statement cannot be checked, but it seems that not all of the information from the available studies was used in the analysis (e.g., feeding rates and timing of life-history events). Also, the author judges the data as 'fit for purpose', which can be questioned as standard ELS tests do not provide sufficient information to parameterize TKTD models, especially as body size is the only sub-lethal endpoint, and that is only determined at the end of the test.

3.3.4 Conceptual/Formal Model

Since this is a highly non-standard 'DEBtox' application, the general evaluation of DEBtox in section 2.3 does not really capture this particular case study. Hence, some model aspects of this case study will be discussed in this evaluation as well. The report (and the paper) provides a detailed summary of the standard DEB animal model, including the relevant equations. The equations for the TKTD module are provided, which is the general module as also used in 'classical' DEBtox (Jager and Zimmer 2012).

The TK equation is the standard one, used in most DEBtox applications. It accounts for effects of growth on TK, but assumes that the organism does not change in shape during growth and that the reserve does not play a role in toxicokinetics. The exchange rate (k_e) is scaled with a surface:volume ratio, and the scaled internal concentration is diluted by body growth. This implies that effects are directly linked to internal concentrations, rather than through some form of damage (as done in GUTS analyses). If effects are driven by some form of internal damage, scaling with surface:volume and dilution by growth are possibly irrelevant (and could lead to bias in the parameter estimates). For eggs, and to a lesser extent yolk-sac fry, the TK model formulation is questionable due to the presence of the yolk sac (using structural body size of the embryo to scale k_e and to calculate growth dilution makes little sense). However, the egg and yolk-sac stages are not considered in this analysis: uptake is assumed to commence when the animals start feeding exogenously. As the author states: "During model calibration of the TK-TD module (§3.1.2), it was found that the effects could only be explained by assuming that exposure only starts after the fish start to feed, which happens approximately at day 43 in the experimental conditions used in the peak-exposure experiment." This modelling choice is supported by the toxicity data in the calibration study (especially the survival pattern) but seems at odds with the remarks in the RAR (page 29-30) where uptake and accumulation in fish eggs were deemed likely. In short, more data points over time would be needed to test the appropriateness of the TK module for this specific case.

For effects on survival, a GUTS-SD module is used (stochastic death), linked to the output of the TK module (DEBtox applications typically do not consider individual tolerance as a death mechanism). For effects on sub-lethal endpoints, a linear-with-threshold relationship between scaled internal concentration and scaled food level (f) is taken. This is standard for this MoA in DEBtox models.

3.3.5 The Environmental Scenario

The 'environmental scenario' used is the one used for the ELS studies (laboratory conditions). From the code, it appears that temperature corrections are made for the relevant parameters. The scaled food level that the organisms experience in the experiments was tuned to make the model reproduce the observed weight at the end of the test. Since the food level required was considerably lower than the presumed *ad libitum* level, and differed between the various experiments (and even between the three cohorts used in the validation study), this would need to be explained. No attempt is made to extrapolate to other than the test conditions. In terms of the life-cycle setting, the analysis only concerns the early life stages: egg, yolk-sac larva and swim-up fry.

3.3.6 Parameter Estimation

Basic parameters for the DEB model (those that govern life history in absence of toxic stress) were directly taken from the add-my-pet (AmP) collection. Entries can be submitted to this collection by anyone, as long as a standardized procedure is followed to extract the parameter values from the available data for the species; a member of the 'board of curators' reviews each entry (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/index.html). The AmP collection is not aimed at regulatory application, but rather to present the best-possible parameter set at this moment for as many species as possible. For some species, the entry is based on a large amount of (good-quality) data, and there can be a large degree of confidence in the appropriateness of the parameter values. However, for others, either less data is available (e.g., several extinct species), or a less-thorough data search was performed. In such cases, the data are supplemented by 'pseudo-data' (based on expected relationships with maximum body size) in the calibration procedure to arrive at a best guess. Therefore, the quality of the entries varies considerably.

The entry for rainbow trout has a completeness mark of 3.5, which is not too high (this mark ranges between 0 and 10, see https://debportal.debtheory.org/docs/Completeness.html). However, as explained in 'Supporting data', the data set used for this particular entry is quite extensive for the early

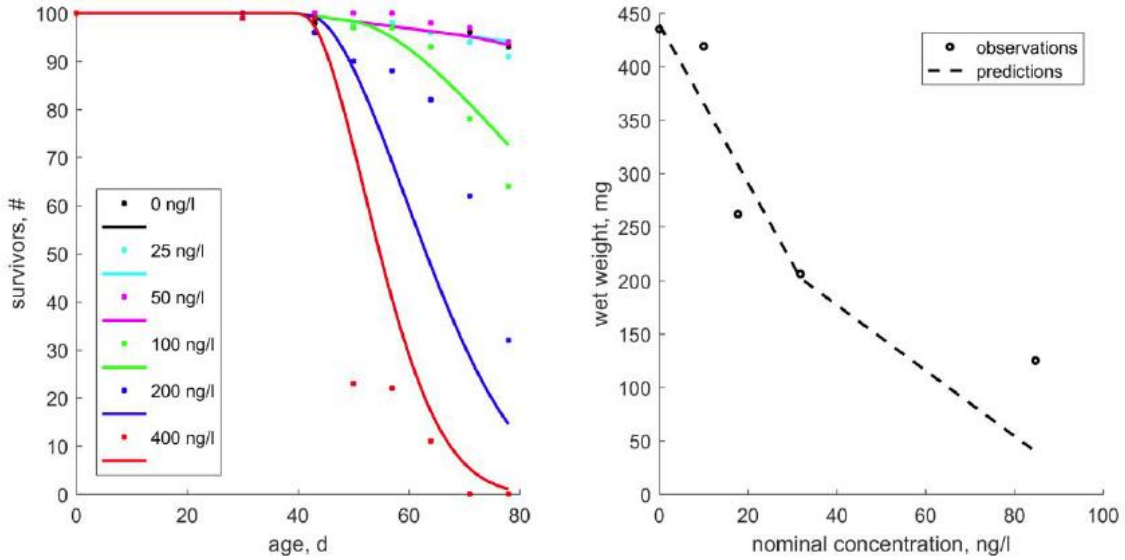
life stages. As the calibrated model provides a good description of the early life history, there can be confidence in its application to toxicant stress as well. However, the parameter values used in the report (dated 27/05/2017) differ from the ones currently on the AmP web site (dated 30/10/2017). For example, the assimilation coefficient differed by 25% and the maturity at birth even by a factor of 6.5. Apparently, the entry has been updated since this analysis was performed; this is confirmed in the paper, where it is discussed to have little influence on the analysis here. It is not uncommon for the standard DEB animal model to be overparameterized, given the availability of relevant experimental data (which implies that different parameterisations will lead to the same fit). Whether this is the case here cannot be judged as the AmP procedure does not provide insight into the parameter uncertainties and correlations (yet). However, a unique parameterization is not necessarily essential for a relevant analysis of toxicity data (although it may become relevant for more substantial extrapolations).

The basic parameters for the species were thus available from the AmP collection (where the scaled food level was set to 1 – *ad libitum* – for all but one data set). However, in the application in the report, the food level was adjusted to make the data match the model predictions for the control. The only data that was available for this purpose was wet weight at the end of the test. The current study thus provides no means whatsoever to test the applicability of the DEB model or the validity of the AmP parameters (which also have changed in the meantime): only one data point is available, and a model parameter is tuned to match it perfectly. The estimated food level for the calibration experiment was very low (0.48) and differed from that established for the validation experiment (0.53-0.73, where a different value was used for the tests with the three life stages). This raises some questions regarding the basic parameterization. More confidence in the results could have been provided by comparing the predictions for time-to-hatch and time-to-swim-up to the observation in the different experiments. Now, the mismatch between the model and the data (in terms of one body-size determination at the end of the test) is completely assigned to a difference in food quantity or quality. While differences in food/level quality between experimental tests are certainly possible, the misfit may have been caused by other factors (e.g., a delay in hatching, relative to the data used in the AmP entry, or differences between varieties of trout). However, the paucity of data precludes any further testing.

Some information is provided on the details of the optimization procedure, although some confusion remains. The author states (Page 39) that the mean relative error is the same as weighted least squares. However, consulting the DEB wiki on the MRE, and comparing it to the weighted least squares used in (Lika et al. 2011), it is clear that they are not the same. Since weighted least squares seems to be the preferred procedure in AmP calculations, it is likely that this is what has been used. The optimization routine makes use of other entries in AmP for related species for initial values. However, a more interesting aspect is that AmP calculations often make use of 'pseudo-data' (conceptually comparable to Bayesian informed prior information). It is unclear whether, and to what extent, these pseudo-data have affected this particular parameterization. No confidence intervals are provided on the parameter estimates, so it is unclear whether the parameters are uniquely identifiable from the data. Settings for the ODE solver can be found in the code.

The plot of the fit on the survival data for the calibration ELS in the code differs considerably from the one presented in the main text of the modelling report. The parameter estimates in the code also do not match the estimates provided in the text and in the paper. It is also interesting that the fit shown in the main text (and the paper) for the control survival over time cannot come from a constant hazard rate (which leads to an exponential decrease in survival probability), unless background mortality only starts at the start of external feeding. Apparently, some model adaptations were made that were not explained. Furthermore, different exposure concentrations are plotted for the two endpoints. For survival, the plot uses nominal concentrations whereas the plot for wet weight shows mean measured ones. The code suggests that the measured values were used in the analysis.

Figure 27: Application of DEB to beta-Cyfluthrin – Fits to the Calibration Data Set



Note that the wet weight in the control (concentration zero) is fitted perfectly as the food level was tuned to match this data point. Also note that the exposure concentrations differ between the two endpoints (likely a difference between nominal and measured concentrations). Effects on mortality only start after t=43, which is the point where the animals start exogeneous feeding (the assumption is made that eggs and yolk-sac fry do not take up the compound). Graphs reproduced from Zimmer (2017).

The author selected feeding (parameter f) as the target process (or MoA) in the analysis. The toxicity data that were used to fit the model provide no clue whatsoever as to the most appropriate MoA. Observations on feeding behaviour apparently support the use of this specific MoA for the model analysis, but results of these measurements are not provided or detailed in the report nor in the paper. The paper refers to Groh *et al*: although they include the wrong paper in the reference list, another paper from Groh *et al* in the same year indeed supports a close link between inhibition of feeding and impairment of growth in fish exposed to pyrethroids.

The fit of the model on the ELS data for constant exposure looks quite reasonable, although the data basis is insufficient for parameterizing (and testing) a TKTD model, especially for the sub-lethal endpoint: only data on body weight at the end of the test, where the apparent food availability f was additionally tweaked to match the control response. No confidence intervals are presented with the model parameters, which implies that it cannot be evaluated whether the parameters can indeed be identified from the data (the model is likely over-parameterised for this data set, as 7 parameters were fitted/tuned to the toxicity data).

It is good to realise that the fit on the survival data depends on the fit to the growth data. In the TK model, body length and growth rate are factors, and these will differ between treatments. Thus, growth influences TK, and thereby the internal concentrations, which affects the survival response. Therefore, it is difficult to judge the goodness of fit for the survival data without some evidence that the DEB model (and the equations for sub-lethal effects) can capture the growth patterns. For sub-lethal effects, the single data point at the end of the test is insufficient to fit or validate the model. In addition, this analysis is based on the assumption that the same TK process underlies lethal and sub-lethal responses (the same k_e is applied to both types of endpoint). Although this is a reasonable assumption, it is possible that lethality is related to a different type of damage than effects through the energy budget. Given the sparsity of sub-lethal effects data, there is no possibility to test this assumption.

3.3.7 Sensitivity and Uncertainty Analysis

No sensitivity or uncertainty analyses were performed, and no confidence intervals on parameter estimates were given. Given the sparsity of the toxicity data, confidence intervals (profile likelihood) would have been helpful to evaluate how well the data can constrain the parameter estimates. The AmP database also does not provide confidence intervals with the parameter estimates (yet), nor is it clear to what extent 'pseudo data' played a role in establishing the entry for this species. A classical sensitivity analysis would not have been particularly useful (see discussion in Jager and Ashauer 2018a).

3.3.8 Comparison with Measurements

The basic DEB model (incl. the AmP parameterization) could not be validated in this study as there is only a single data point (wet weight at the end of the test in the control), and the model was actually calibrated to match that data point. The AmP parameterization for rainbow trout is based on a rather extensive data set, including data for the early life stages, so there can be some confidence in this parameter set. Nevertheless, parameters may differ between varieties of the same species (especially for rainbow trout), and possibly depend on test conditions. Therefore, there would be considerably more confidence in the validity of the standard model and the AmP parameterization if they were demonstrated to match the control response in the ELS study (on more aspects than just the body size at the end of the test). A close match of the model to the control data is essential to provide a meaningful mechanistic interpretation of the toxic effect.

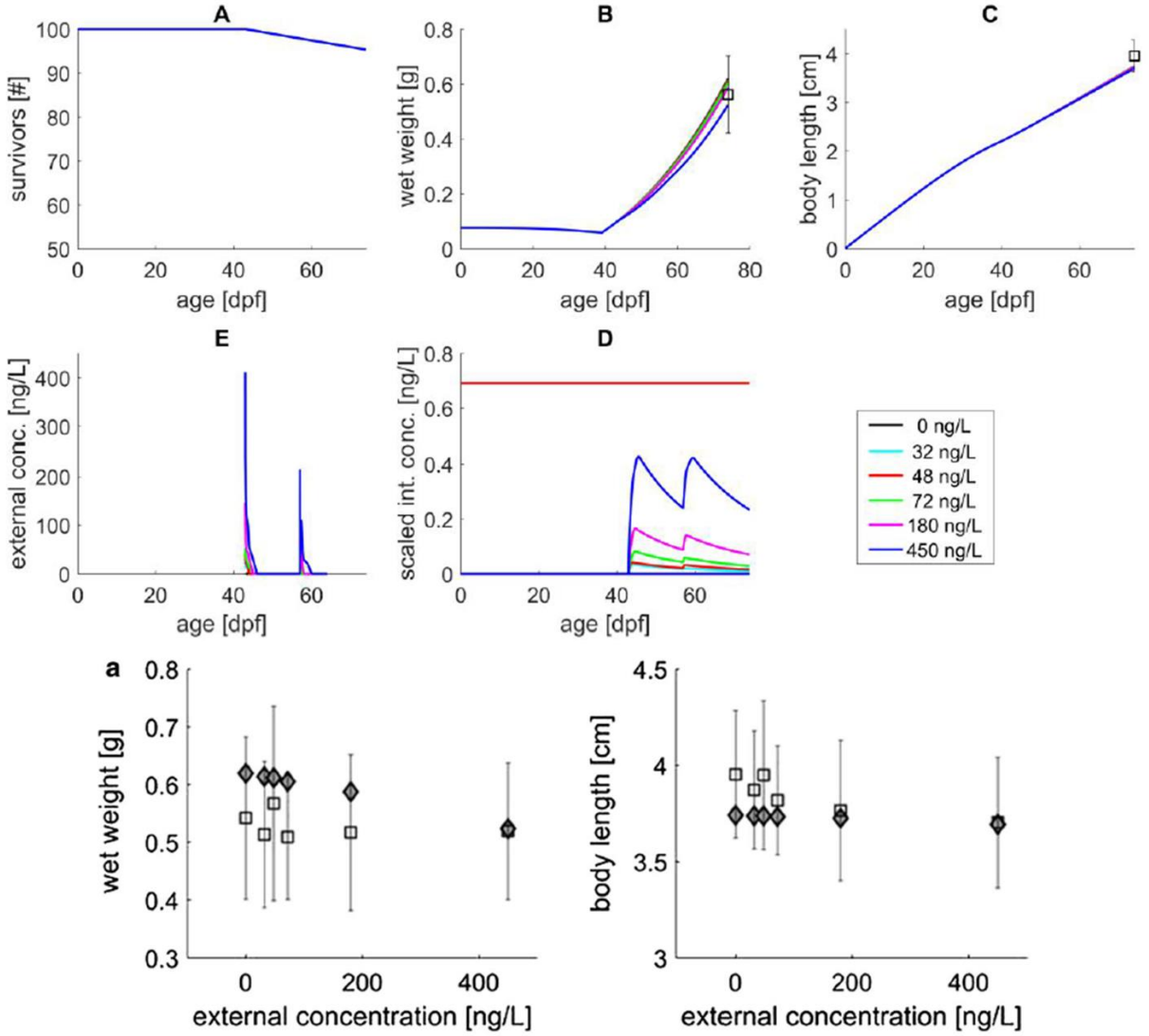
The complete model (with TKTD module) was calibrated on one data set and used to predict independent data from a second experiment. However, also in the validation case, a calibration was used as the scaled food level was matched to the observed weight at the end of the test in the control for each tested life stage (Table 6). Like the calibration study, the validation study is also extremely limited, with observations of body size at the end of the test only (other endpoints were available but not used, see section on 'Supporting data' above). There is almost no effect in the validation study, for none of the cohorts. This was consistent with the model prediction, which implies that the model is indeed able to explain the difference between the two studies. However, it provides no support for the validity of the model, nor for its ability to predict toxic effects under other conditions.

An indication of the validity of the basic parameterization was obtained by comparing the parameter values for the species of interest to parameter values established for related species (section 7.7.1.1. of the modelling report). This indicated that the parameter set is not unreasonable. The author also states that the curators of AmP will perform a check on the 'implied properties' of the parameterization. However, the results from this check are not available. Nevertheless, the existence of such an independent check is laudable and helps to provide standardization and a level of quality control over the AmP collection.

In relation to the validation results (and supported by the plots for scaled internal concentration), the author states that "The model shows that beta-cyfluthrin has fast dynamics (rapid uptake and elimination)." However, this does not correspond to the extremely low value for k_e established in the calibration (Table 4 of the modelling report: 0.0001283 d^{-1}), which implies a depuration half-life of 15 years. This discrepancy may be partly explained from the scaling of k_e with body size (it is unclear which maximum size was used in this scaling). Toxicokinetic depuration is, however, not the only process affecting the time course of the internal concentration; it is possible that growth dilution explains the rather rapid responses observed. For the validation graph (reproduced here as Figure 2), we can see that the elimination of the compound after the pulse is not extremely rapid: less than 50% of the estimated scaled-internal concentration is eliminated after some 11 days. From the increase in body weight over time, we can roughly estimate that growth dilution would lead to an 'elimination' half-life

of 10 days, very close to the observed elimination. It is therefore possible that toxicokinetic elimination is very slow (and k_e is indeed very low), but that the decrease in body residues (model predictions as shown in the validation plots) should be assigned to growth dilution. It is good to realise that growth dilution depends on the growth rate; under different conditions (when growth is impaired by the toxicant or by food limitation), growth dilution will be of less importance, and the 'elimination' of the compound after a pulse may be substantially less (and hence prolonged toxic effects may occur).

Figure 28: Application of DEB to beta-Cyfluthrin – Model Validation



Results for the model validation for the cohort exposed as swim-up larvae. Swim-up larvae were most sensitive in the calibration test. Top panel shows the model predictions over time. Only the data point for the control is shown as the treatments fell within the standard deviation of the control. Graphs reproduced from Zimmer (2017). Lower panel shows the results at the end of the test versus concentration (grey diamonds are model predictions). Graphs reproduced from Zimmer et al. (2018, edited).

3.3.9 Model Use

The author states that "The model shows that results from both studies (Carlisle 1985 and Ramsden 2017) are consistent." This is a reasonable conclusion, supported by the presented analysis. They continue: "Therefore, the model can be used in the context of ecological risk assessment for beta-cyfluthrin" and "Thus, we can use the model to make predictions for untested exposure scenarios …" These conclusions are, however, not supported by the presented analysis. The problem with the current analysis is not so much with the model (although a number of modelling issues were flagged in this evaluation), as with the data: the data from the ELS studies is insufficient to calibrate and validate the model (even with the help of the AmP collection for the basic parameters, and even if there would have been a substantial effect in the validation test). More size measurements over time, and preferably also additional endpoints (e.g., estimation of yolk content, feeding rates), are essential to support the model analysis. Also, the lack of relevant case studies (with more extensive data) for other compounds/species of this particular type of model application (toxicity to early life stages) makes it hard to place confidence in the current analysis.

The approach taken in this application (using standardised and evaluated parameter sets from the add-my-pet collection, and the standard DEB model) may be a good strategy for routine application of DEB-based models in ERA in the future. Especially because of the structure, organization and parameter collection that are already in place. Toxicity tests will never provide sufficient data to parameterize the standard DEB model (although they *can* for simplified DEB model versions, which would be a different strategy to consider for application in ERA), so making use of reference parameter sets from AmP makes a lot of sense. However, this particular application to early-life stages would need to be supported by a few case studies with more extensive data sets. The downside of this strategy is that the standard DEB model is relatively complex and not very transparent, and that the quality of the entries in the AmP collection varies considerably between species (and is difficult to judge). A quantification and propagation of parameter uncertainties (and demonstrating that parameters can be identified from the data), both for the basic parameters and the toxicity parameters, would help provide confidence in the analysis (though such an error propagation across an optimization will be technically challenging).

3.3.10 Overall Judgement

In this particular dossier, the model analysis was only added as additional information, and was not used for extrapolation. The model was able to explain the difference between the two ELS studies (constant versus pulsed exposure), but calibration and validation would be insufficient to support extrapolations of effects to other conditions. Even though the model application to ELS studies is promising, more work is needed on the model structure and testing (e.g., TK for eggs and yolk-sac larvae), and some case studies with more extensive data sets (this application of 'DEBtox' is quite novel). The main problem is that ELS standard protocols are not suitable for calibrating and validating a TKTD model, so an extension of the test design beyond the protocol seems essential for a proper model application. In this particular study, an additional limitation is that the validation study showed no effects and hence provides little support for the model. Therefore, extrapolations made with this model and this parameter set would be unsupported at the moment.

The current evaluation also flagged several issues with the model analysis such as a lack of confidence intervals, different fits in the code section and the main section, feeding rates were apparently measured but not provided or used, potential differences between cyfluthrin and beta-cyfluthrin, and use of tests with and without sediment.

3.4 IDamP – Application to Bromoxynil-octanoate

Evaluation by Tjalling Jager

3.4.1 Background

IDamP (Preuss et al. 2009a) is an individual-based population model (IBM), not spatially explicit, for *Daphnia magna*. The original paper only deals with the basic life history of *D. magna* in the absence of toxicants; only food stress and crowding effects were considered. Later papers include modules for toxicants. In the study of Preuss (2015)[30] for the risk assessment of bromoxynil-octanoate, the model is not used to simulate population dynamics, but to simulate 21-day reproduction studies for constant and time-varying exposure. This is not a typical application of this model, and the conclusions from this evaluation therefore should not be transferred to the use of IDamP in general.

It should be noted that IDamP does not include consideration of toxicokinetics (or damage dynamics) for sub-lethal effects: a time-independent relationship between exposure concentration and level of effect on a target process (here feeding rate) is assumed. Therefore, it does not classify as a TKTD model. However, it can provide a good approximation for chemical-species combinations where effect dynamics are closely linked to the external concentration (i.e., fast toxicokinetics and fast damage dynamics, which implies rapid onset of effects and rapid recovery).

3.4.2 Problem Definition

In standard *Daphnia* reproduction tests, a (roughly) constant exposure concentration is provided to the organisms over 21 days. The compound under consideration is rapidly degraded in the environment, and only short (less than 1 day) exposure peaks are expected. Application of a 21-day NOEC or ECx to the peak concentration of a pulsed exposure scenario would therefore very likely produce conservative results. The author refers to the EFSA Aquatic Guidance Document (EFSA PPR 2013) to support the use of TKTD modelling to assess the effects due to short, pulsed, exposures (although IDamP does not qualify as a TKTD model, as explained above). From the conclusion of the modelling report (Preuss, 2015), it becomes clear that the main reason to apply the modelling was to make a case for the use of a time-weighted average exposure concentration (PECtwa) rather than the peak concentration (PECmax) for pulsed exposure.

3.4.3 Supporting Data

A 21-day reproduction study with *D. magna* under constant (flow-through) exposure was used to calibrate the model. The study report for this test (referred to as Putt 1991) was not available for this evaluation, but details are provided in the modelling report of Preuss (2015). This study included regular observations of offspring production, but body length was only determined at the end of the test. Due to the absence of data on body length over time, the data set is thus limited for the purpose of model calibration, and limited for evaluating the appropriateness of the model and of the selected mechanism of action (inhibition of feeding).

The calibrated effects model was tested with an independent data set for *D. magna* under time-varying exposure (Kent et al. 1993, report also not provided but details in the modelling report). This was also a 21-day reproduction test, but the setup was static renewal: the medium was refreshed 3 times a week, and the compound could not be detected just prior to the next renewal. This implies rapid disappearance of the substance under these conditions but does not allow calculation of the degradation rate (and thus precludes accurate definition of the exposure scenario in this study; only a maximum DT50 can be established).

[30] Modelling report: T.G. Preuss (2015). Virtual pulse exposure reproduction study of Bromoxynil-octanoate on *Daphnia magna* with the IDamP model. Report No.: EnSa-14-1051. Document No.: M-497765-02-1. Confidential Report.

For the calibration, measured concentrations were used, converted to octanoate equivalents. For the validation data set, the situation is unclear. The legend of Figure 7-2 from the modelling report (see Fig. 30 below, top panel) mentions measured concentrations (supported by the small differences between the data points in the graph) whereas the text and Table 7-2 mention nominal ones (possibly, the initial measured and nominal concentrations were very close in this case).

Table B.9.4.4-20 in the Draft Renewal Assessment Report (DRAR, ANSES 2017, Vol.3, B9, CP, p. 135-136) shows that survival was somewhat reduced in the highest three treatments of the calibration data set (only significantly in the highest treatment). Effects on survival were not included into the model analysis, and it is unclear whether and how the reproduction data are corrected for the death of certain individuals. According to OECD test guideline for *Daphnia* reproduction, only offspring from surviving mothers should be included in the analysis (although for dynamic modelling, it is best to include all offspring). However, the level of mortality is not very high (15% in the highest treatment). For the validation data set, no information on survival was available for this evaluation.

The IDamP model itself contains a parameterization for the basic *D. magna* life history, derived from extensive experimental data from the University of Aachen (see also the general evaluation for IDamP in section 2.4).

3.4.4 The Environmental Scenario

The model was first compared to the control data from the flow-through test. In the modelling report (Preuss, 2015), the author states that "the environmental scenarios of the experiments were mimicked by the model as far as possible (as far as given in the report). This scenario was tested and parameterized using controls only." The following model parameters were adapted to match the situation of the toxicity test: number of daphnids per beaker, volume of test beaker, flow speed, and temperature. The food situation could not be exactly matched as the toxicity test employed a combination of algae and fish food (possibly because the compound in question is a herbicide, and thus to avoid indirect effects via the food of the daphnids). For the model, both the algal concentration and the maximum size of the daphnids were tuned to (visually) match the observed data in the control. The animals in the control of the toxicity test grew to a larger size and had a larger reproductive output than can be achieved with IDamP under the default settings. This modification is acceptable as fish food was provided in the test, next to algae, whereas IDamP's default parameterization considers algae as food only (fish food is likely of higher nutritional value and may lead to larger ultimate body size). Furthermore, different clones of *D. magna* are known to differ somewhat in life history (the clone used is not mentioned).

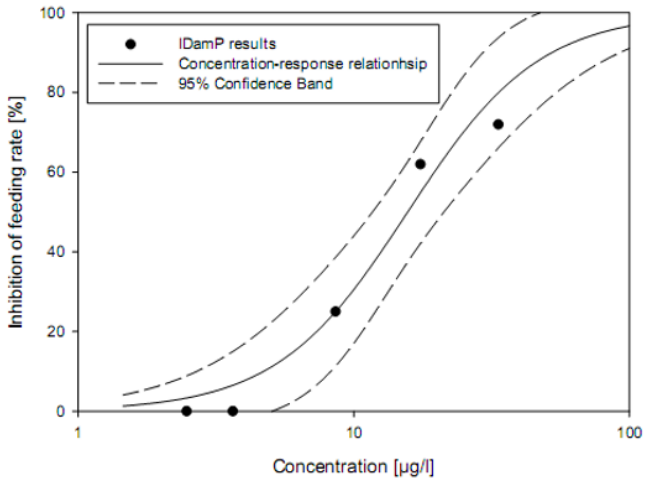
Also for the validation study, the environmental conditions in the model were set as to mimic the conditions of the experimental test (same parameters modified as for the calibration). In this case, the animals had been fed with algae and no adjustments to maximum size were made to match the control data (although the food concentration was indicated to be set by 'expert judgement').

3.4.5 Parameter Estimation

The model was first compared to the control data in the flow-through test. As described in the previous section, the environmental conditions for the model were matched to the test conditions; some manual tuning was needed for the food concentration and the maximum size of the daphnids, based on the observations in the control. Given this tuning, it is perhaps not surprising to see a nice correspondence between predicted and observed length and reproduction in the control. However, the fact that the model is capable of the simultaneous prediction of both endpoints, and manages to capture the pattern of reproduction over time, support the usefulness of the model. Given that the model has been extensively validated already under a range of conditions (Preuss et al. 2009a), this provides confidence that the model is indeed appropriate for the traits under control conditions.

Toxic effects on sublethal endpoints are included in IDamP as a static dose response curve, linking the external concentration directly to a certain model parameter. Hence, the model does not include a TK module for sublethal effects (and no 'damage dynamics'). In this case, the model parameter for which a dose response was used was the feeding rate. No rationale is provided for why this should be the most representative mechanism of action, apart from the statement that "Due its mode of action it can be assumed that bromoxynil-octanoate reduces the amount of energy in the daphnid which is available for growth and reproduction." The compound is a herbicide, but there is no reason why herbicides should always affect feeding capabilities of daphnids. Herbicides will affect the algae provided as food in the test, but the reproduction test design should not rely on the algae being alive and growing (besides, fish food was supplied to the daphnids as well). If a different effect mechanism can explain the observed effects as well, it is unlikely to severely bias the results in this study (there is no extrapolation to other food conditions or to the population dynamics, for which selecting the correct mechanism will be much more important). Nevertheless, additional analyses for different mechanisms would have strengthened the conclusions.

Figure 29: Application of IDamP to Bromoxynil-octanoate – Fitting of Individual-Level Effects



Dose-response curve as used in the model analysis. Inhibition percentages resulted from tuning IDamP to yield a close correspondence with the actual effect patterns. Graph reproduced from Preuss (2015).

The calibration of the dose-response curve is rather awkward. IDamP is run with various inhibition percentages of the feeding rate, and it is determined at which inhibition percentage the patterns from the model match the observations for length (at the end of the test only) and reproductive output (fitting by eye). On these manually-calibrated inhibition percentages, a log-logistic dose-response curve is fitted (as reproduced above). Figure 6-3 in Preuss (2015), p. 12 subsequently compares the predictions from the model (based on the predicted feeding inhibition from the dose-response curve) to the observations for the various treatments. This rather awkward procedure was probably followed because the IDamP implementation does not include routines for optimization. The calibrations themselves are not provided, so the 'fitting-by-eye' cannot be checked. This is unfortunate as it complicates evaluation of the appropriateness of the selected mechanism of action (which is already heavily compromised by the fact that body length is only determined at the end of the test). Confidence intervals are plotted on the dose response curve, but these are not very meaningful given the fact that the inhibition percentages were fitted by eye. Even though the calibration is rather qualitative, it provides a reasonable (and generally worst-case) representation of cumulative reproduction of individuals under constant exposure (as demonstrated in Fig. 6-3 of the modelling report).

3.4.6 Sensitivity and Uncertainty Analysis

The fitted dose-response curve is shown with 95% confidence bands, but, since it is fitted on points that were themselves calibrated by eye, these intervals have limited usefulness and they are not propagated to model predictions. Otherwise, no parameter uncertainties are quantified. Another source of uncertainty, inter-individual variability, *is* addressed. A range of simulations are performed where variability between individuals is included stochastically and propagated to the model output. At birth, each individual receives a set of independent random factors modifying maximum filtration rate, growth rate, juvenile development rate, embryonic development rate, brood size, and expected lifetime (based on experimental data, but all included as independent variabilities). The bands produced by this analysis (min and max of 3000 simulations) are very wide (Fig. 6-1, 6-3, 7-1 in Preuss 2015) and probably exaggerate the impact of variability between individual daphnids.

No sensitivity analysis was performed, but such studies would not have yielded much useful information in this context anyway (Jager and Ashauer 2018a).

3.4.7 Comparison with Measurements

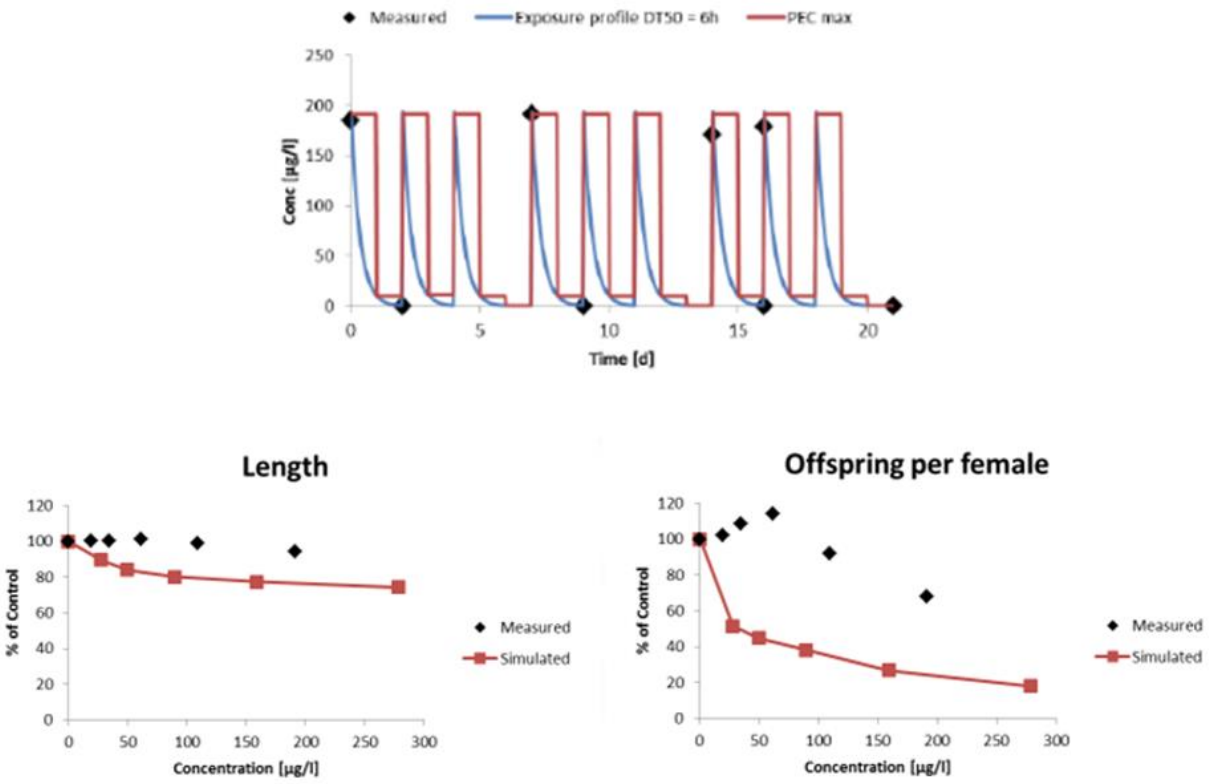
The calibrated effects model was tested with an independent data set under time-varying exposure (also 21-day reproduction with *D. magna*). Figure 7-1 (in Preuss 2015, p. 13) compares control response predicted by the model to that observed in the data set (unlike for the calibration data, the model was not tweaked to match the control data, apart from setting the algal concentration). Body length at the end of the test was predicted well, but the reproduction data were largely underpredicted by the model, which may again relate to differences in food quality or differences between labs and/or clones.

The setup of this experiment was static renewal, with the medium renewed three times a week. The compound could not be detected at the end of the renewal interval (48 or 72 hours). The author states that “a DT50 of 6 h was used to calculate the exposure concentrations over time”, later followed by “because the temporal resolution of the IDamP model for exposure and effects is one day the maximum concentration per day were taken as exposure scenario.” Figure 7-2 (Preuss 2015, reproduced as Fig. 30 below) suggests that what was used in the modelling is a pulsed scenario, with a constant exposure scenario over one day, determined by the calculated initial (maximum) concentration on that day. This can be considered a highly worst-case representation of the actual exposure of the daphnids.

In the comparison between model and data for the treatments, only the results at the end of the test (21 days, both body length and cumulative reproduction) are shown (Fig. 7-3, reproduced as Fig. 30 below). A comparison of the time responses (as in Fig. 7-1) would have been much more useful to evaluate model performance (especially as it could allow testing the assumption of rapid TK/damage dynamics). In the comparison, the points for the model predictions are plotted at different concentrations than the measured data; concentrations are here nominal concentrations. It is unclear why the model was simulated for different exposure concentrations than those used in the experimental test (see x-axis position of the points in Fig. 30 below). The model does predict considerably stronger effects than observed in the study, which may be explained from several worst-case choices in the analysis. The main reason is probably the worst-case (and unrealistic) definition of the exposure scenario for this analysis: the assumption of constant exposure during one day at the maximum level, whereas, in reality, the concentration will decrease rapidly after a renewal. It is perhaps counter-intuitive, but assuming a *worst-case* exposure scenario in a validation comparison, as forcing of the model, is actually *best-case* for the model performance in ERA (the model will look more conservative than it actually is). Thus, the fact that the model produces overly-conservative predictions when fed with an overly-conservative exposure scenario is not very relevant; it says nothing about the level of conservatism in the model or in the model calibration. This validation therefore does not help to judge the

model's validity and predictive power, nor to establish its degree of conservatism. A more realistic validation in this study is, however, hampered by the lack of concentration measurements within a shorter distance from the renewal event (precluding estimation of a degradation rate), the lack of body-length observations at intermediate time points, and the minimum time step of one day in IDamP. Nevertheless, using a time-varying exposure with a reasonable DT50 (i.e., 6 hours), and a smaller time step in the model, would already provide a far more meaningful validation exercise than using the block-pulse scenario.

Figure 30: Application of IDamP to Bromoxynil-octanoate – Model Validation



Validation of predicted bromoxynil-octanoate effects for a different exposure scenario. Assuming an unrealistically worst-case exposure scenario (red line in top plot) leads to worst-case predictions for the effects (lower plots). Graphs reproduced from Preuss (2015).

What the validation data set does show is that pulsed exposure, up to rather high peak concentrations (that show strong effects when applied as constant exposure), does not lead to strong effects on growth and reproduction. Therefore, effects are not related to the maximum exposure concentration in this test. However, this provides no information on whether TK/damage dynamics is fast (and a static dose-response curve is reasonable) or slow. For sub-lethal effects, as long as first-order kinetics for TK/damage dynamics applies, time-varying exposure should always produce less effect than constant exposure at the same PECmax (over the same time window). A proper TKTD analysis would be needed to test/demonstrate the speed of TK/damage dynamics. Showing the entire validation data set, with reproduction over time, could also be possibly used to support the use of the static dose-response curve.

3.4.8 Model Use

The calibrated model was used in a series of simulations. First, the 'most-sensitive time window' was established by predicting the effects after 21 days as a result of a 3-day exposure event starting at different times in the total 21-day simulation (Fig. 8-1, p. 16, in Preuss, 2015). Day 1-3 was the most sensitive window, and this (day 1) was subsequently also used for other pulse durations (1, 2, 3 and 7 days). This sensitivity to early exposure probably reflects the importance of slower growth, and thereby a delayed start of reproduction, which does not occur when exposure starts at a later age. These simulations are, strictly speaking, insufficient as the sensitive window may be different for a different pulse duration. However, this is unlikely to seriously affect the model results for cumulative reproduction due to the use of a static dose-response (also reflected by the almost constant effect percentage for the different start times of exposure in Fig. 8-1). Next, EC10s were calculated for different pulse durations. It is good to realize that although the validation study indicated substantial overprediction of the effects by the calibrated model, these EC10 are *not* necessarily worst case. The validation overpredicted the observed effects as it was fed with an unrealistically worst-case exposure profile. The simulated EC10s are calculated for a constant concentration during a pulse, and for such an exposure scenario, the model predictions are not expected to be specifically conservative. The claim for 'conservative predictions' therefore does not hold for these EC10s.

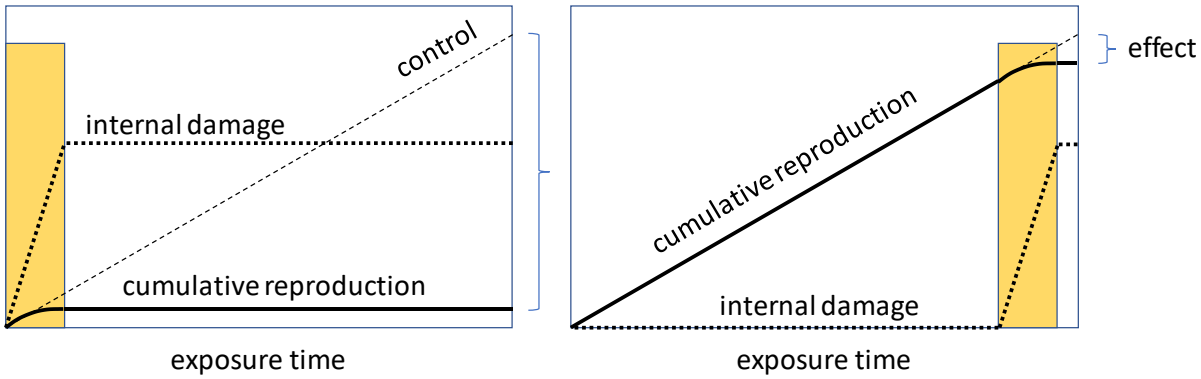
The author claims "linear reciprocity of exposure time and concentration" for reproduction: doubling the exposure duration (within a fixed total time period) leads to roughly half the EC10. First, it needs to be clarified that this is not the regular interpretation of the EC10 (which can only be meaningfully defined for constant exposure), and also not the regular interpretation of Haber's rule (the 4-d EC10 should be half the value of the 2-d EC10, both after constant exposure). In IDamP, a static dose-response curve is used, and hence the EC*x* for cumulative reproduction (under constant exposure) can only remain constant over time. For this reason, Haber's rule (which assumes that EC*x* decreases with increasing exposure duration) cannot result from IDamP, and the model cannot yield linear reciprocity. The fact that EC10 does not stay constant in the IDamP simulations is related to the fact that this metric is here defined in a particular manner, as the concentration that, when applied as a pulse of *t* days, will lead to 10% reduction after 21 days. This definition is perhaps more in line with the idea of using a PECtwa for risk assessment purposes, but not in line with Haber's rule predicting that effects should become stronger with duration of (constant) exposure.

A major problem in this respect is that Haber's rule has no theoretical underpinning. It can follow mathematically from TKTD models, but only under very special (and biologically unrealistic) conditions. Furthermore, its relationship to the use of PECtwa is tenuous. Basically, this relationship only holds for irreversible quantal effects such as mortality (and only when Haber's rule applies over the entire time window over which the PECtwa is calculated). For sub-lethal endpoints, such as cumulative reproduction, it would be exceptional to see Haber's rule apply over any length of exposure time; EC*x* values do not generally smoothly decrease with time, and the time pattern would depend on the endpoint and how it is expressed (see e.g., Jager 2011).

To illustrate the failing of Haber's rule i.r.t. PECtwa, a simple thought experiment might help. Assume that we follow a fully-grown adult organism that is reproducing at a constant rate (and we follow cumulative reproduction as endpoint). If Haber's rule applies, the EC*x* should decrease over time under constant exposure. In a situation of fast kinetics, and fast damage dynamics, the internal damage level that determines the effect is constant and a static dose-response relationship applies. A constant reduction of the reproduction rate implies a constant effect on cumulative reproduction, relative to the control, and hence a constant EC*x* over time (i.e., Haber's rule does not apply). A decreasing EC*x* over time requires a slow increase of the internal damage level over time. In fact, it can be shown that Haber's rule can mathematically emerge from such a simple model for infinitely slow damage dynamics, in the absence of a threshold for effects, and with a specific functional form for the relation between damage and effect on the reproduction rate. However, for very slow damage dynamics, PECtwa is not a

good descriptor for effects. This is illustrated in Figure 31, where both exposure profiles have the same TWA concentration. A pulse early in the time window leads to a strong effect on cumulative reproduction: slow kinetics prevents recovery after the pulse, so the effect (relative to the control) will grow ever larger over time. A pulse late in the time window will have little effect as the animal will have reproduced at control rate for most of the time window. This illustrates that the length and position of the time window will have a strong effect on the perceived effect (even though the effect is in essence the same in both cases). Furthermore, it seems inappropriate to 'reward' a situation of no, or very slow, recovery (and thus prolonged effects upon short exposure, and cumulation of effects over multiple exposure events) by use of a PECtwa (which is obviously lower than PECmax).

Figure 31: Application of IDamP to Bromoxynil-octanoate – Exposure Profile vs. Effect Size



Influence of exposure profile on observed effect size. Example of two exposure profiles with the same TWA, but a very different effect under slow kinetics/dynamics (slowness is a prerequisite for Haber's rule to emerge under constant exposure). The control (thin broken line) is reproducing at a constant rate. Graph provided by Tjalling Jager.

Clearly, the link between Haber's rule and applicability of PECtwa does not hold (at least not for cumulative reproduction), and its degree of (un)protectiveness depends on the exposure pattern and the time window over which the concentration is averaged. Furthermore, it is good to note that use of PECmax will not always be worst case; underprediction of effects may occur when the EC*x* is derived from a test that is short relative to the exposure pattern from which PECmax is derived, and when this exposure patterns contains many (or prolonged) exposure events.

For sub-lethal effects, the fact that a substance-species combination does not show Haber's rule therefore does not mean that use of PECtwa will lead to poor predictions (as illustrated by the IDamP analysis), and vice versa. Unfortunately, the guidance on the use of PECtwa (EFSA PPR 2013) is rather confused. It is important to stress that the use of PECtwa (or PECmax) with an EC*x* (or NOEC) is also a model, and an extremely crude and poorly tested one. This model may provide a reasonable (or worst-case) representation of effects in some cases, but it is unknown what these cases are without detailed experimental testing (or model simulation). Thus, the RMS comments on the "ratio between the acute EC50 and the NOEC based on mortality" and the requirement of "increased exposure duration results in increasing sub-lethal effects" are irrelevant for the applicability of PECtwa (apart from the fact that these are of course the requirements put forward by EFSA).

After this critique on the link between reciprocity and PECtwa, as assumed in the EFSA guidance, it is good to focus on whether the PECtwa is applicable in this case (even though Haber's rule is not). The author makes the case for using PECtwa by showing that the product of EC10 and pulse duration is rather constant. However, it is important to note that the author's conclusion on "linear reciprocity of exposure time and concentration" is based on model simulations. That the model has this behaviour is interesting; it is likely a consequence of the (unsupported) assumption of fast toxicokinetics and damage kinetics. However, since it is not demonstrated that the model provides a useful (or even worst

case) representation for time-varying exposure, the model analysis provides no support for the applicability of PECtwa.

In this case, the validation data set under semi-static exposure clearly shows that much higher concentrations can be tolerated when the organisms are exposed in pulses compared to constant exposure: under constant exposure, 33.5 µg/L gave strong effects on reproduction, whereas in the validation study, this (as peak concentration) was still in the range of no effects. This indicates that the use of a PECmax, with an ECx for constant exposure, will tend to produce worst-case results (at least for this time window, under these test conditions). Would the TWA concentration from the test be a better descriptor for effects in the validation data set? If the DT50 of 6 h (as assumed in relation to the IDamP simulations) is reasonable, a quick calculation can shed some light. For a treatment with three pulses per week at 191 µg/L, reproduction was roughly 70% of the control. Assuming exponential decay and a DT50 of 6 h, this treatment has a TWA of 29 µg/L. For the treatment with constant exposure, the closest exposure level (33.5 µg/L) reduced reproduction to 20% of the control (and hence a much stronger effect). A constant exposure to 8.6 µg/L yielded a reproduction of close to 70% of the control; to yield this TWA concentration with pulses of 191 µg/L would require a DT50 of less than 2 h. If we can exclude such short half-lives for this compound, this quick and dirty calculation indicates that a TWA would be on the safe side in this case (for this time window, these test conditions, these types of pulses, and as long as both toxicity studies are comparable in other respects than the exposure profile). However, that is no guarantee that the same conclusion will hold for other exposure profiles (e.g., FOCUS output).

It is good to stress that this model analysis focusses solely on sub-lethal effects. However, effects on survival (immobility) occur in a similar range of concentrations (the LOEP reports a 2-day EC50 of 26 µg a.s./L). Sub-lethal effects are reversible (at least in the model), but mortality is not. Therefore, in extrapolation to pulsed exposure with high peaks, short duration, it would be good to consider effects on mortality as well.

3.4.9 Overall Judgement

The problem to which the model is applied (link between constant and pulsed exposure) requires a TKTD modelling approach. However, the IDamP model is not particularly suitable for TKTD analyses; it is an individual-based population model that is used here to mimic individuals in a 21-day test. Firstly, its application as a TKTD model is hampered by the fact that it employs a time step of 1 day, which is crude given the rapid degradation rate of the compound in this study. More pertinently, IDamP does not qualify as a TKTD models since it excludes TK as well as damage dynamics (it employs a static dose-response between external concentration and the feeding rate). The use of a static dose-response relationship implies that toxic effects closely follow the exposure concentration. This overestimates the effects caused by an exposure peak but also overestimates recovery after a peak. It implies that the model is only suitable for chemicals that have rapid toxicokinetics and damage dynamics (which means that the damage in the organism will closely follow the changes in external concentration). This might be the case here, as the calibration data indicate a rapid onset of effects and no increase in effect strength, but complete and rapid recovery would still need to be demonstrated.

The data sets used for calibration and validation of the model are limited due to the lack of observations on body size at intermediate time points, and the lack of information on the disappearance rate of the compound in the validation study. Overall, it is unclear what the model analysis is adding on top of the two data sets that are available. The validation analysis showed that the calibrated model will provide overly-conservative predictions for pulse-exposure scenarios, when the model was run with a clearly unrealistic worst-case exposure scenario. However, this is a rather trivial result, and does not help to provide any support for the general applicability of the model for more realistic exposure scenarios (and no guarantee that extrapolations will be conservative). The analysis suggests that consid-

erably higher EC10 values can be used for short-term pulsed exposure than for 21-day constant exposure, but this conclusions rests on the assumption in the model of rapid and instantaneous recovery after a pulse. The validation data set, however, clearly shows that higher maximum concentrations can be tolerated when provided as pulses rather than constant concentration, and suggests that a TWA concentration would be on the safe side (for this particular experimental test situation, although the actual exposure pattern in this test was unknown). Since the TWA concentration heavily depends on the time window over which the averaging takes place, this does not translate into support for using a PECtwa for any exposure profile and time window.

The whole section on 'linear reciprocity' is troublesome, also because the guidance on this aspect is confused (EFSA PPR 2013). Haber's rule is a rule-of-thumb from an era when there were no computers and little mechanistic knowledge on toxicity. This rule has no mechanistic basis, though it may provide a reasonable approximation for some compounds, in some species, for some endpoints, for some exposure durations (and some concentration ranges). Furthermore, for sub-lethal endpoints (in any case for cumulative reproduction), there is no link between the usefulness of Haber's rule and applicability of PECtwa. The whole point of TKTD modelling, however, is to get rid of this awkward juggling with descriptive rules-of-thumb in dealing with the time aspect of toxicity. Using TKTD models (even though IDamP does not qualify as one) to prove or disprove such rules is putting the cart before the horse. In almost all practical cases, the response to a toxicant will depend on the actual profile of exposure; profiles with the same PECtwa can give rise to different, and sometimes extremely different, toxic effects. Therefore, the only sound solution to the problem of time-varying exposure is to apply calibrated (and validated) TKTD models to make predictions for the effects given the actual profiles or a specific window from such a profile (see also EFSA PPR 2018).

3.5 IDamP – Application to Pirimicarb

Evaluation by Tjalling Jager

3.5.1 Background

This evaluation for the application of IDamP in the risk assessment of pirimicarb is based on a modelling report of Preuss et al. (2009b)[31]. Additionally, the EFSA conclusions on pirimicarb (EFSA 2005)[32] were consulted for details on data that were used in the modelling report. IDamP (Preuss et al. 2009a) is an individual-based population model (IBM), not spatially explicit, for *Daphnia magna* (see evaluation of the general model in section 2.4). The original publication only deals with the basic life history of *D. magna* in the absence of toxicants; only food stress and crowding effects were considered. Later papers included modules for toxicants.

In the case study of Preuss et al. (2009b), the *Daphnia* life-history traits were made temperature dependent, which is explained in detail in the appendix of the modelling report. This extension was more recently published in the open literature (Gabsi and Preuss 2014). The original IDamP model focuses on population development in a homogeneous laboratory environment. In this case study, the model is used in connection to a lake model (StoLaM) to provide a more realistic environmental setting for the Daphnids. StoLaM predicts temperature, light and nutrient conditions, and from that calculates algal population development. IDamP uses the output of StoLaM as inputs for the Daphnia population development, and feeds back the number of Daphnia to StoLaM (which takes care of the grazing pressure on the algae). The combined model was termed DaLaM and was also published recently (Strauss et al. 2017).

3.5.2 Problem Definition

The compound in question is an insecticide from the carbamate family (AChE-inhibitor) for which reversible effects might be expected. Daphnids turn out to be the most sensitive aquatic organisms (*D. magna* and *D. pulex* were at least one order of magnitude more sensitive than a range of other tested aquatic species; based on acute and chronic statistics such as L/EC50 and NOEC). Therefore, the focus on Daphnia seems warranted. The purpose of the modelling study was to simulate the impacts on Daphnia populations following time-varying exposure (according to relevant FOCUS exposure profiles). Due to the specific sensitivity of Daphnids, effects on individuals are almost inevitable. However, the explicit aim was to investigate recovery at the population level.

3.5.3 Supporting Data

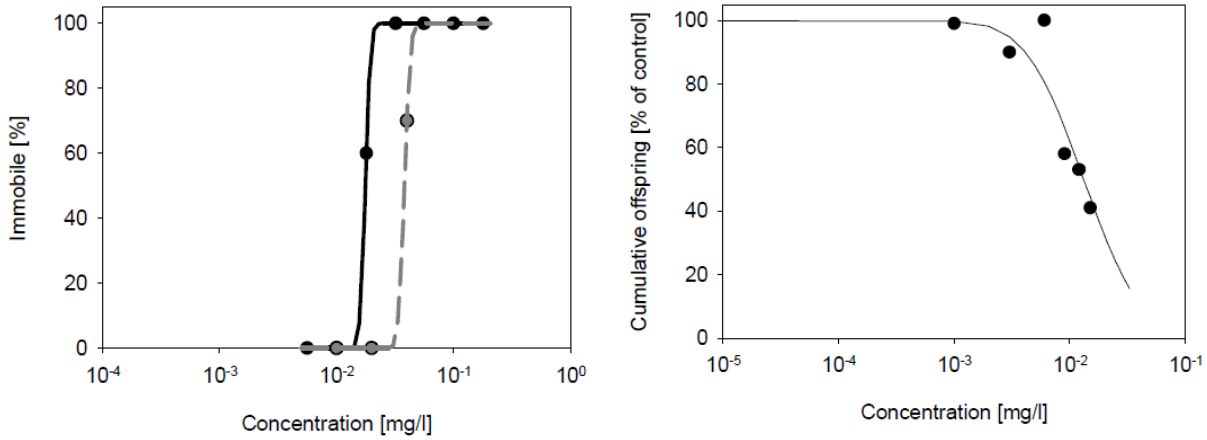
Several acute toxicity studies with *Daphnia* are mentioned in the modelling report of Preuss et al. (2009b). Apparently, for acute toxicity, a standard study of Kent & Shillabeer (1996) was used (data not provided for this evaluation) for neonates, and a non-standard study from the open literature (Kusk 1996). Kusk reports the results from several experiments, a 48-hour water-only toxicity test with neonates and 6-7 day old juveniles, and a 72-hour accumulation/toxicity study for adults, with added sediment (with survival, reproduction and body weight reported). However, the modelling report is not entirely clear regarding which data sets are used: Fig. 6-7 (reproduced in Fig. 1 below) is said to compare 48-h dose-response data for neonates and adults from Kusk (1996), but that paper does not provide the raw data for neonates (this could be results from Kent & Shillabeer, 1996), and

[31] Preuss, T. G., T. Strauss, M. Hammers-Wirtz and H. T. Ratte (2010): Modelling effects of pirimicarb on Daphnia magna populations under simulated field conditions. Research Institute for Ecosystem Analysis and Assessment - gaiac; Institute for Environmental Research (Biology V), RWTH Aachen, Worringerweg 1, D- 52056 Aachen, Germany. Report No. T009915-06. Sponsor: Syngenta, Jealott's Hill, Bracknell, United Kingdom. Sponsor Ref. No. PP62_10059. Confidential Report.

[32] EFSA (2005): Conclusion regarding the peer review of the pesticide risk assessment of the active substance pirimicarb. EFSA Scientific Report 43: 1-76.

the test with adults is for 72-hours in the presence of sediment. From Fig. 6-7 it appears that both survival data sets only have one concentration with partial survival, which implies large uncertainty in the slope of the dose-response curve.

Figure 32: Application of IDamP to Pirimicarb – Fitting of Individual-Level Effects



The left panel shows the acute toxicity data used to derive the dose-response curve. The black dots are for neonates (from a standard test) and the grey dots for adults (Kusk 1996; non-standard test with added sediment). The right panel shows the results from the non-standard reproduction study (10-day, static, with added sediment). Effects on growth (not shown) based on a 3-day study with sediment. Graphs reproduced from Preuss et al. (2009b).

This results in two dose-response curves being used for effects on survival (mobility) in the model (see Fig. 32): one for neonates (EC50 = 18 µg/L) and for adults (EC50 = 38 µg/L). This seems to indicate that adults are less sensitive than neonates. However, Kusk (1996) concludes that older animals are just as sensitive as newborns (comparing the 48h-EC50s for water-only exposure of neonates and 6-7 day-old animals). The 72-hour test with adults yielded a higher EC50 as the test was different, especially due to the addition of sediment (also according to Kusk). Sediment was shown to reduce accumulation and effects of the toxicant, presumably due to dissolved organic compounds released from the sediment, which reduced bioavailability. This implies that the dose-response relationships used for neonates and adults in the modelling study cannot be directly compared. The statement in the modelling report that neonates are more sensitive than adults is thus not supported by the results of Kusk (in fact, 6-7 day old juveniles were, if anything, *more* sensitive than neonates). The different results for mortality are shown in the Table below; only the first and last value were used in the modelling study.

Table 17: Application of IDamP to Pirimicarb – Acute Sensitivity of *D. magna*

Study	Life stage	Exposure	Duration	EC50
Kent & Shillabeer	neonates	Water only, unknown whether nominal or measured conc.	48 hour	18 µg/L
Kusk	neonates	Water only, nominal conc.	48 hour	21 µg/L
Kusk	juveniles (6-7 d)	Water only, nominal conc.	48 hour	16 µg/L
Kusk	adults	Added sediment, nominal conc.	72 hour	38 µg/L

From Preuss et al. (2009b, edited), as derived from data in Kusk (1996) and in Kent & Shillabeer (1996).

The results of Kusk (1996) were based on nominal concentrations. However, Kusk demonstrated that 93-106% of the compound remained in the water phase after 72 hours. Therefore, it can be assumed that the nominal water concentrations are representative for the actual exposure.

The 3-day study of Kusk with adults (with added sediment) was also used to quantify effects on growth (as dry weight) for the modelling. However, since the animals were not fed in this test, the results on body weight are difficult to interpret; the observed differences in weight between the treatment likely reflect different rates of shrinking, and are biased by the fact that body weight will include contributions from the reproductive buffer and/or the eggs in the brood pouch. Therefore, this study cannot be used to derive a dose-response curve for growth to be used in population modelling.

For effects on reproduction, a study from Hamer & Goggin (1997, not provided) was used. The EC50 for reproduction was in the same range as the acute EC50 for immobility (13 µg/L). The EFSA conclusion on pirimicarb (EFSA 2005) clarifies that this is a 10-day static study with added sediment. A 21-day semi-static study without sediment resulted in a considerably lower NOEC (0.9 µg/L versus a NOEC of 6 µg/L in Hamer & Goggin). The 21-day study is likely the one made available for this evaluation (Thompson et al. 1989), which is not referred to in the modelling report of Preuss et al. (2009b). The effects in this data set are stronger than the effects in the 10-day reproduction study that was used for the modelling (see table below). No reasons for selecting the 10-day test are given. The presence of sediment has likely reduced bioavailability of the compound in this test as was discussed in relation to the acute toxicity tests (total measured water concentrations would overestimate the bioavailable dissolved concentration that likely drives toxicity if dissolved organic carbon is released from sediment).

On Page 36 of the modelling report, it is mentioned that OECD standard tests were used to derive the dose-response relationships. However, this is only true for the acute effects on neonates (according to the information on Page 27-30). The acute effects on adults, and the chronic effects on growth and reproduction were taken from non-standard tests with non-standard duration and added sediment. Given that the model was set up to mimic a small pond, tests with added sediment may be more representative (the EFSA conclusion on pirimicarb also states that the "RMS proposes to use a study in the presence of sediment"). However, the degree of conservatism will depend on whether the water concentrations reported by the FOCUS calculations represent total-dissolved or freely-dissolved aqueous concentrations. Clearly, more scrutiny of bioavailability is needed to judge to what extent the concentrations from the toxicity tests can be compared to those derived from the FOCUS calculations.

Table 18: Application of IDamP to Pirimicarb – Chronic Sensitivity of *D. magna*

Study	Exposure	Duration	NOEC	EC50
Hamer & Goggin (1997)	Added sediment, static, unclear whether nominal or measured conc.	10 day	6 µg/L	>15 µg/L (repro, EFSA) 13 µg/L (repro, modelling report)
(Thompson et al. 1989)	Water only, semi-static, measured concentrations	21 day	0.9 µg/L (length) 1.7 µg/L (repro)	1.7-3.6 µg/L (repro) 2.5 µg/L (survival)
Kusk (1996)	Added sediment, static, no food, nominal concentrations	3 day		58 µg/L (body dry weight, modelling report)

From Preuss et al. (2009b, edited).

Several validation data sets were selected: mesocosm data from Aachen University. Several studies without toxicant (untreated control mesocosms, to study the natural patters of daphnid population dynamics), and only one study with an insecticide (not the compound of the dossier). It is not explained why these specific studies were selected.

3.5.4 The Environmental Scenario

StoLaM was set up to mimic the conditions in a mesocosm pond (4.9 m^3, 1m depth). "The StoLaM sub-model was used to simulate realistic dynamic environmental conditions as they occur under field conditions." A minimum temperature of 4°C was set, due to uncertainties in biological parameters for even lower temperatures. Data from a meteorological station at Aachen University were used to provide a realistic setting. The outputs for temperature and algal density were fed into IDamP, which was ran for a smaller sub-volume of 10 L. One- and two-year simulations were conducted. There is some confusion over the time step, which is reported as 5 minutes on Page 12 and 1 minute on Page 13 of the modelling report. However, IDamP runs in discrete steps of 1 day (Page 13), which is rather coarse (especially given the high temporal resolution in the StoLaM part).

The food web in the model system is a very simple one: nutrients -> one algal species -> Daphnia (and back to nutrients). Thus, there are no predators or competitors for the daphnids. Since the populations are reduced to low numbers as a result of the pesticide application, the absence of predators and competitors is a critical element of the model analysis. The only stressors considered for the Daphnids are toxicant stress (from the compound of interest, due to the application) and effects of food limitation and crowding at high densities.

For the model simulations for the compound of interest, two environmental settings were used: one mimicking a laboratory setup (standard IDamP: 1 L beaker, 20°C, 0.5 mg C in food per day), and one mimicking a pond mesocosm (linked to StoLaM) as explained above. The chemical exposure pattern is applied according to several FOCUS scenarios: "The exposure was calculated using FOCUS Surface Water scenario using the maximum total carbamate residue (parent and metabolites containing the active carbamate moiety) on a given day." The exposure profile is thus a worst-case approximation of the actual profile. However, in this model setting, this does not lead to a worst case for the toxicity. Due to the assumption of a static dose response without recovery from the effect, the actual exposure pattern has become irrelevant; only the height of the peak concentration matters (a 1-second pulse will have the same effect on an individual as a 1-day pulse). For the laboratory system, the 1-day block pulses are applied when the population is in steady state (total population size is kept in check by crowding effects and/or food limitations). For the mesocosm setting, the environmental scenario for a specific year was selected, which led to low population abundances in the controls (assumed to be a worst-case situation).

The IDamP model was extended with effects of temperature. The Arrhenius relationship was used, although the equation given is incorrect, or at best incomplete: the exponent was missing. This exponent was included in earlier work (Gabsi and Preuss 2014), so this is probably an error in the text of the modelling report (Preuss et al. 2009b). If this did not provide a good description of the data set, another model was used (various descriptive curves were fitted). In general, this procedure is consistent with the rest of IDamP: many independent equations that provide a good description of each life-history traits (under a single changing environmental factor). The mechanistic basis is weak, but this is not necessarily problematic for model application (as demonstrated by the, in general, good performance in the various validation studies).

3.5.5 Parameter Estimation

For the parameters governing the Daphnia life history (in absence of toxic stress) the values derived in the basic IDamP implementation are used (Preuss et al. 2009a); that parameterization is discussed in the general evaluation of the model (section 2.4). Static dose-response curves are used in IDamP to include the toxic effects on different endpoints of the daphnids. In general, it is good to stress that the problems flagged for the data sets selected (see section 'Supporting data') also affect the parameterization.

Mortality/Immobility

The dose-response for immobility after 48 or 72 hours is used, assuming immediate effect at this level and no recovery from immobility. Given that the pulse duration from the FOCUS profile is 24 hours or less, and that recovery is possible (Andersen et al. 2006), this seems like a worst-case assumption (although immobility might equal death under natural conditions, e.g., due to increased predation risk). The individual-tolerance (IT) concept is applied (Page 29), which means that at the LC50, after 48 hours, the 50% of the animals that are immobile are the sensitive ones, and the insensitive ones remain. It is not entirely clear how the distribution of sensitivities is implemented here; the fraction effect from the dose-response relationship is compared to a random number from a uniform distribution that each individual draws at birth (Page 29). This is probably correct, but IT is not necessarily worst case for more than one pulse: as the individuals surviving the first pulse are the tolerant ones, they will also survive the second pulse (due to the lack of TKTD processes in IDamP, carry-over toxicity is impossible in the model). The competing explanation of 'stochastic death' (SD) would predict the same fraction reduction in the population after the first and the second pulse, which would lead to stronger effects in case of multiple pulses. The EFSA Scientific opinion on TKTD modelling (EFSA PPR 2018) therefore suggests to run model predictions for both IT and SD, and generally use the worst case for risk assessment.

Reproduction

For effects on reproduction, the results from a static 10-day study (with added sediment) were used, assuming that the dose-response at the end of the test can be used as the immediate effect on reproduction that the organisms experience. Furthermore, no recovery of reproduction is included: once exposed, the individual will remain inhibited for its entire life time. In general, that would be a worst-case setting, although it is not clear how rapidly the compound disappeared in the static toxicity test, and what concentration was used for the x-axis of the dose-response relationship (i.e., nominal, initial measured, time-weighted average).

Growth

For effects on growth, the results from (Kusk 1996) were used again: a 3-day bioaccumulation/toxicity test with adults, with addition of sediment. Also here, the dose-response was used as an immediate effect on growth without the possibility for the individual to recover. Kusk measured body *weight*, but IDamP works with body *length*. It is unclear if a conversion is made (implementing the effect percentage observed on weight on the length of the daphnids would lead to worst-case results: 10% reduction in weight implies only 3.5% effect on length, assuming that the animals have the same shape as in the control). However, as explained under 'Supporting data', this experiment cannot be used to calibrate a dose-response curve for effects on growth, as the animals are not growing (they are not fed and hence will be shrinking).

Level of Conservatism

As the effect model assumes no recovery of individuals, it is generally worst case with respect to the exposed animals. However, it assumes implicitly that the neonates born from exposed mothers are behaving like the control animals (no carry-over or trans-generational effects considered) and hence ensure rapid recovery of the population. From the toxicity tests presented, this assumption cannot be evaluated. However, the results of Andersen et al. (2006) show that exposure to a single short pulse of the compound as a neonate leads to a reduction in the size (and hence reduced fitness) of their offspring later in life. Therefore, there seems to be some possibility of delayed and prolonged effects that is not considered in the model, and not represented in the toxicity data used for model calibration/parameterisation.

3.5.6 Sensitivity and Uncertainty Analysis

No sensitivity or uncertainty analysis was performed. Also, the uncertainty in the dose-response relationships (i.e., the confidence band on the effect level) was not quantified or propagated. This would have been especially interesting for immobility as the dose-response curves only have one treatment with partial effects (which should lead to substantial uncertainty in the curve, especially at the tails). Variation in life history between individuals was assumed and propagated to the population output. At birth, each individual receives a set of independent random factors modifying maximum filtration rate, growth rate, juvenile development rate, embryonic development rate, brood size, and expected lifetime (based on experimental data, but all included as independent variabilities). Also, for the toxicant effect, each individual receives a threshold for effects at birth, derived from the dose-response relationship for mortality (although it is not clarified how the assumed difference between neonates and adult sensitivity was implemented). The dose-response relationship is thus viewed as the distribution of thresholds in the population, which is consistent with the IT view of mortality (but not the SD view).

The authors state that (Page 33): "Recovery was defined as the average of the treatment population reaching within or above the 95% confidence interval of the control population, i.e. that the difference between treatment and control was not statistically different." However, it is incorrect to treat the intervals as 95% CIs: they represent variation in population simulations, but including variation between individuals only, and it is not entirely clear how this variation between individuals is quantified from experimental data (see the general evaluation of IDamP in section 2.4). Furthermore, the variation was included independently on the various traits (the random value an individual receives for trait A is unaffected by its random value for trait B, and vice versa; in reality, traits could well be correlated). Therefore, these results cannot be interpreted in terms of statistical significance. It is good to note that this criterion implies that the average affected population should recover to the lower edge of the 'confidence band' for the control, which is not a worst-case definition. Nevertheless, some pragmatic definition of 'recovery' is needed as both the control and affected populations will vary dynamically over time (in a different way).

3.5.7 Comparison with Measurements

The basic IDamP and StoLaM models have been validated quite extensively; in the appendix of the modelling report, a summary is provided of the most relevant comparisons of model predictions to independent data. The temperature module of IDamP was tested at the population level with unpublished laboratory data. Correspondence between predicted and observed population density was good; in the modelling report, only result for an untreated laboratory population was shown (Fig. 7-1), though IDamP has also been validated against data for toxicant-exposed laboratory populations (see the general evaluation of IDamP in section 2.4).

In Chapter 8 of the modelling report, validation results for DaLaM on mesocosm studies from Aachen University are provided. These studies used three species of Daphnid, but the sum was compared to the simulated population dynamics for *Daphnia magna* as used in the model. DaLaM was compared to results for mesocosm studies. The model was parameterized using the nutrient (phosphorus release rate) and weather conditions for the actual mesocosms, but the authors state that no calibration was performed. The authors compared the modelled versus the measured water temperature in the mesocosms, which indicates that the model is successful in predicting this environmental factor. The population dynamics of the daphnids in the control mesocosm was reasonably well captured by the model (Fig. 8-3, especially as no calibration was applied). Especially the spring peak in abundance was well captured; the summer peak that occurred in several years was underestimated (Fig. 8-3, which the authors claim represents a worst-case aspect of the model: small populations are at higher risk of extinction). Given the relative simplicity of the model, and the ecological complexity of the mesocosm, the predictions are very reasonable.

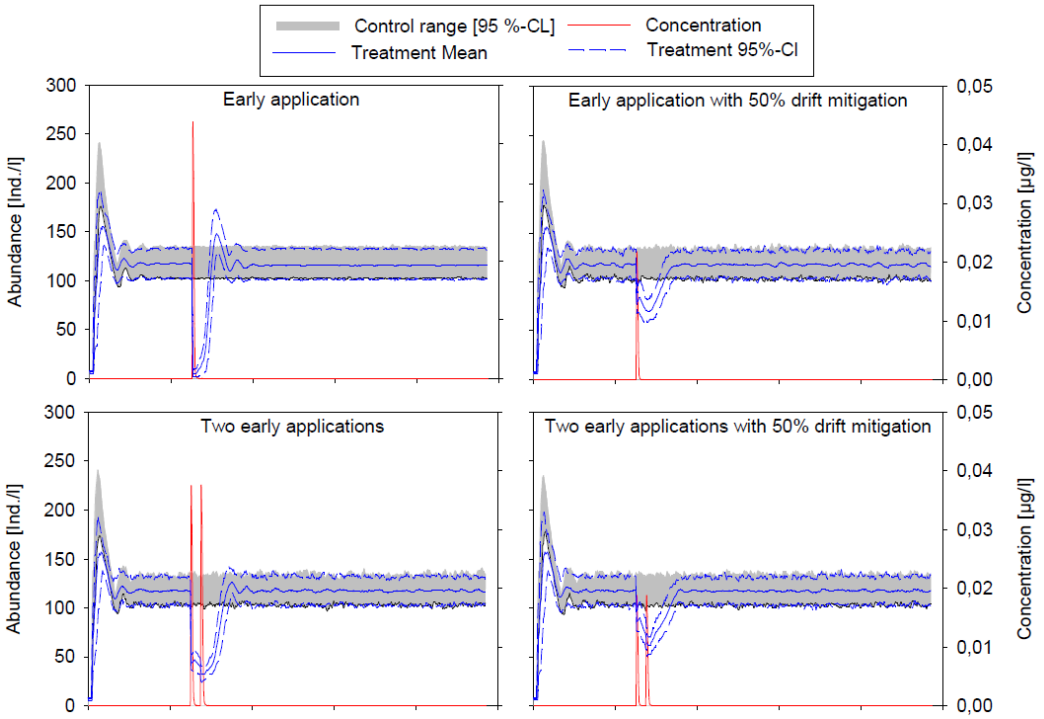
Also, a comparison for the algal populations (Fig. 8-5, chlorophyll density) was made. Even though the observations started too late in the year to fully catch the spring bloom, the predictions seem to be in the right range.

Finally, a comparison is made between observed and predicted population dynamics for a mesocosm exposed to an insecticide (Fig. 8-6, not the chemical of the dossier). The chemical was assumed to instantly kill 99% of the Daphnids at each day of application, and be completely dissipated after 1 day, but it is unclear on what basis this specific scenario was used: it seems like the 99% mortality is based on the observed effects in the mesocosm, which would then imply that the prediction is not completely independent. Again, given the simplicity of the model as a whole, relative to the complexity of the system, the prediction of the population dynamics for daphnids (and to a somewhat lesser degree for algae) is quite convincing. Nevertheless, without more information on the individual-level effects of the compound (in laboratory toxicity tests and as implemented in the model used to make the predictions), it is difficult to generalize these findings. No experimental studies at the population level have been performed with the compound of interest.

3.5.8 Model Use

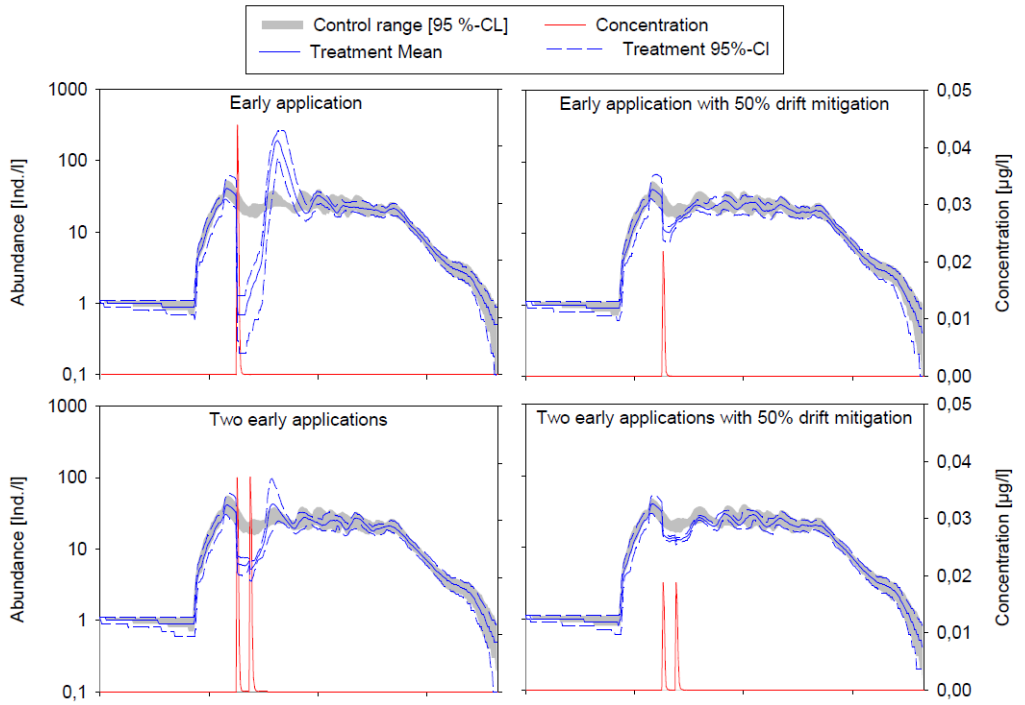
Simulations for the insecticide under scrutiny were conducted for a laboratory setting (standard IDamP) and for a field situation (mesocosm pond as simulated with DaLaM). The following endpoints were calculated from the simulations: population dynamics, maximum effect and time to recovery. The various exposure scenarios had a short-term effect on the population dynamics. Considerable mortality occurred (especially for the single ‘early application’; nearly wiping out the population), followed by recovery. There is no recovery assumed for individuals; the recovery of the population occurs as (presumably) fit neonates (in all respects comparable to those in the control) are produced by the surviving mother. After severe population reduction, these neonates will have plenty of food in the model as grazing pressure is low, and hence are capable of supporting rapid exponential population growth. Note that in the two-pulse scenarios, the second pulse does not have an effect on population reduction anymore. This relates to the assumption of IT made in the model (an SD assumption could also have been used, applying x% effect for every exposure pulse), and the lack of TKTD processes (precluding carry-over effects).

Figure 33: Application of IDamP to Pirimicarb – Simulations for Laboratory Scenario



Model simulations for a laboratory setting with different application scenarios. Concentration axis shows units of µg/L but this should have been mg/L. Graphs reproduced from Preuss et al. (2009b).

Figure 34: Application of IDamP to Pirimicarb – Simulations for Pond Scenario



Model simulations for a mesocosm-pond setting with different application scenarios. Concentration axis shows units of µg/L but this should have been mg/L. Note that abundance is here plotted on log scale. Graphs reproduced from Preuss et al. (2009b).

3.5.9 Overall Judgement

Overall, it is likely that the DaLaM model is capable of providing a realistic picture of *D. magna* population dynamics in representative field situations (those where influence of competition with other species and predation are negligible). The coupling of the two models is logical, and substantial validation efforts have been performed, both for the individual models and for the combined one.

The toxicity module (part of IDamP) lacks toxicokinetics and damage dynamics; instead, a static dose-response curve is used. The effect occurs immediately after exposure and was here assumed to be irreversible. This will generally be a highly worst-case assumption for the individuals. Recovery from immobility due to the compound was demonstrated (Andersen et al. 2006), although it was not complete and, in a natural setting, immobility may imply *de facto* death. The use of IT here, and the lack of TKTD processes, nevertheless implies a best-case assumption for multiple pulses as both choices preclude carry-over toxicity to the second pulse. IT makes sure that there are only insensitive individuals left after the first pulse, and the lack of TKTD ensures that there is no body residue or damage left after the first pulse to which a second pulse can add.

There is no information on recovery for sub-lethal effects. However, the results of Andersen et al. (2006) show that short pulse exposure (to high concentration) early in life had lasting effects throughout a 21-day reproduction test. A rather best-case assumption in the model is that neonates born from exposed mothers are assumed to be clean and to behave like control neonates. This is a best-case assumption as there may be maternal transfer (chemical transport from mother to egg) or maternal effects (damage to neonates born from exposed mothers) that would have gone unnoticed in the toxicity tests. The results of Andersen et al. (2006) hint at that possibility as the neonates born from mothers (themselves exposed as neonates) were of a somewhat smaller size.

The toxicity module is parameterized with data sets that cannot be considered worst case, especially due to the addition of sediment in most of the tests that were selected for parameterization. Sediment addition was demonstrated to decrease bioavailability and reduce the toxic effects (Kusk 1996). However, it is good to note that the EFSA conclusion on pirimicarb states: “To refine this risk the RMS proposes to use a study in the presence of sediment.” Furthermore, a best-case situation is created at the population level as there are no predators or competitors in the model, nor additional stressors such as parasites or other toxicants.

The result of the simulations should not be very surprising: in the absence of predators, competitors and other stresses, Daphnia populations can quickly rebound, even from severe reductions in population size, as long as healthy neonates (with the same fitness as in the controls) are produced and sufficient food is available. Therefore, there is actually very little information gained from the modelling study, and the regulatory question could be rephrased to: how certain are we that sufficient animals will indeed survive the toxic insult, and how certain are we that healthy neonates are produced from exposed mothers? The latter question was to some extent answered (Andersen et al. 2006): when neonates are exposed to a short pulse at high concentration, when they start reproducing as adults, there are (limited) delayed effects on their offspring’s fitness. However, the small effects on offspring body weight (5-17%, not related to pulse duration, only one dose tested) may not be very relevant, biologically. The first question can be further divided into two parts: the uncertainty in the dose-response relationship, and what ‘sufficient’ means. These are discussed separately below.

Uncertainty in the Dose-Response Relationship

The uncertainty in the dose-response relationship was not included into the analysis (i.e., no confidence bands on the dose response, and no propagation of this uncertainty). Since the dose-response relationships, used for model parameterization of mortality, were very steep (only one dose with partial effects), the uncertainty in the tails will be very high. At a predicted effect level of 94% for the la-

boratory simulation (which is in the extreme tail of the sensitivity distribution), we should not be surprised to see 100% effect, simply due to the uncertainty in the dose-response relationship. Furthermore, the only reason that there was some survival in the worst application case is that the dose-response relationship for adults was based on a test with added sediment. Using the results for neonates or 1-week old animals, tested without sediment, would have produced 100% mortality in this scenario. The bioavailability reduction by sediment was likely caused by the release of dissolved organic compounds. Such compounds are also present in natural waters (especially the ponds as used for the modelling). However, such bioavailability aspects need to carefully considered in all parts of the analysis (also with regard to the FOCUS profiles used for the model predictions).

An additional source of uncertainty in the dose-response relationship is the fact that the laboratory test was done with a single clone under rather optimal conditions (apart from the lack of food in the survival tests). Even though the effect of temperature on the life history was included in the model, temperature may also interact with the toxic response, which was not included (there are also probably no data to base such interactions on). Furthermore, the presence of other stressors may lead to more effect from the compound of interest in a field situation than predicted from the dose-response.

Clearly, there is a considerable and unquantified uncertainty in the dose-response relationship for the laboratory cohort, and additional uncertainty when using this relationship for field populations (exposed to multiple stressors). Therefore, some margin of safety seems appropriate.

What is Sufficient Survival?

Owing to the settings of the population model in DaLaM, very few individuals (perhaps even one) would suffice to allow recovery of the population (apart from the situation at the end of the year when also the control population will be in decline). In real systems, this is unrealistic. When the Daphnia populations will be severely reduced, the addition of other species (as competitors, predators or parasites) might push them over the edge of extinction. In this respect, it is noteworthy to mention the authors' discussion on the substantial impact that backswimmers had on population density in one of the mesocosms used for validation.

As already discussed, the toxicity module assumes that surviving mothers will produce healthy offspring. In the study of Andersen et al. (2006), animals received a much higher dose than the animals will experience in the FOCUS profiles, but for a much shorter duration (several hours). Clearly, surviving neonates grew to adulthood and reproduced offspring, albeit less and of a slightly inferior body size. It thus appears that there are delayed effects, and possibly slow or only partial recovery of individuals after pulse exposure. I therefore do not fully agree with the author's statement on the lack of "latency" (Page 57). Whether these second-generation neonates would still show adverse effects is unclear (but probably unlikely).

Furthermore, the toxicity module assumes that the survivors of the first pulse are the less sensitive ones (IT), and that the survivors recover from exposure immediately (no TKTD). Therefore, subsequent pulses will have no additional effect on individuals in the model. Andersen et al. (2006) showed that there is recovery after a pulse, but that a second pulse after a recovery period of 48 hours will lead to increased effects on immobility (and mortality). Clearly, 48 hours between exposures is insufficient to warrant against carry-over effects between the exposure events. For the FOCUS scenarios, a 1-week interval between applications was used. It is unclear whether that would be sufficient for the individuals to recover. The authors also support their assumption of lack of "latency" (i.e., delayed effects) with: "additionally no mortality was observed in a 21 day reproduction test at constant exposure up to 15 µg/l (Kusk 1996)." However, (Kusk 1996) performed a 3-day test, and saw no mortality at 20 µg/L (with added sediment, in absence of food).

Exactly what level of toxicant-induced mortality would still be acceptable to ensure successful population recovery is unclear. In any case, some margin of safety again seems appropriate. A level of 96%

mortality, even if there would be no uncertainty in that number, seems insufficient to safeguard the population from a substantial extinction risk.

Overall Conclusions

Even though DaLaM is an impressive attempt to model Daphnia population dynamics, the model analysis in this case study has a number of drawbacks: the awkward selection of toxicity tests for parameterization, the lack of TKTD elements in IDamP, the sole focus on IT as death mechanism, an incomplete documentation, and the lack of additional stressors such as predation and competition. These drawbacks do not so much reflect the models themselves (e.g., a GUTS module has been implemented in IDamP as well as predation/competition modules) but rather the choices made for this particular case study.

Perhaps more pertinent is that the analysis performed for this dossier provides little useful additional information for the risk assessment. In a system with only nutrients, algae and daphnids, the daphnids will always be able to quickly recover from even very severe population reductions. The crucial questions in this case (how certain are we that sufficient individuals survive, and will the survivors produce viable offspring) are not served by the insights from population and hydrodynamic modelling. They require more detailed TKTD considerations, and IDamP is poorly equipped for those. A straightforward analysis with a TKTD model for survival (GUTS, SD and IT), supported with a toxicity test on the viability of offspring from survivors of a pulse, would have provided more useful insights. The modelling report is inconsistent and unclear in places, and the drawbacks listed above lead to substantial uncertainties in the model simulations. Even without these specific drawbacks, some margin of safety seems prudent if Daphnia populations are to be safeguarded. Especially the scenario with the ‘single early application’ seems to be cutting it very thin (especially since abundance is plotted on log scale for DaLaM).

3.6 IBM *Chaoborus* Population Model – Application to beta-Cyfluthrin

Evaluation by Jeremias Becker

3.6.1 Background

The IBM *Chaoborus* population model has been applied for the risk assessment of several active substances. Here a case study from the risk assessment of the active substance beta-cyfluthrin was evaluated that has been provided by the German Federal Environmental Agency, UBA, which acted as Rapporteur Member State. Beta-cyfluthrin was first approved by the European Commission for the use as insecticide on 1st January 2004 (European Union 2019). In the frame of a regular review of the registration (Annex I Renewal), the environmental risk of spray applications of this pyrethroid insecticide to growing crops of cereals and potatoes was re-evaluated in 2017. In the review, beta-cyfluthrin failed to pass Tier 1 and Tier 2 of the revised EU risk assessment framework (EFSA PPR 2013) because of its high toxicity to aquatic macroinvertebrates. Larvae of *Chaoborus* sp. were identified as the most sensitive taxon in a microcosm study. Therefore, the applicant proposed a modelling study of Strauss and Norman (2017)[33] using the Chaoborus IBM population model to demonstrate that populations of *Chaoborus* sp. may recover from exposure in ponds through reproduction and recolonization.

3.6.2 Problem Definition

Experimental studies identified *Chaoborus* sp. as the freshwater macroinvertebrate taxon being most sensitive to cyfluthrin or beta-cyfluthrin (Strauss and Norman 2017). The LOEC for *Chaoborus sp.* larvae (1.6 ng/L) was lower than the maximum predicted environmental concentration for surface waters (PEC_{sw} = 3 ng/L) according to FOCUS step 4 modelling. If *Chaoborus sp.* was not considered, an Ecological Threshold Option Regulatory Acceptable Concentration (ETO-RAC) for aquatic organisms of 2.5 ng a.s./L could be derived from the mesocosm study, based on a NOEC of 5 ng/L for the 2nd most sensitive taxon *Crangonyx pseudogracilis* and an assessment factor of 2. Assuming that further mitigation measures will be applied, the PEC_{SW} may be reduced to meet this ETO-RAC that excludes the most sensitive taxon *Chaoborus sp*.

Therefore, Strauss and Norman (2017) assessed the potential of *Chaoborus sp.* populations to recover under realistic field conditions in a simulation study using the *Chaoborus* IBM population model (version 4.01). The aim was to inform the allocation of an Ecological Recovery Option (ERO-) RAC for the affected taxon *Chaoborus* sp. The ERO-RAC may allow temporary (weeks to months) effects at the population level (EFSA PPR 2013). By demonstrating that *Chaoborus sp.,* can recover from pesticide effects within this time period, the authors suggested to exclude this taxon from the calculation of the ETO-RAC. Using this ERO-RAC of 2.5 ng/L, beta-cyfluthrin may have passed the risk assessment with additional mitigation measures.

3.6.3 Supporting Data

The IBM *Chaoborus* population model (Strauss et al. 2016) was developed based on a comprehensive data set on the life history of the model species *Chaoborus crystallinus*. The origin of parameter values is well documented and a number of experiments have been described that were performed to parameterize species-specific parameters of the model. See the general model description in section 2.6 for more details.

Microcosm studies of Heimbach (2000) on cyfluthrin, evaluated in Hommen and Heimbach (2000), and of Jenkins (2014) on beta-cyfluthrin indicated that *Chaoborus* sp. is the most sensitive taxon

[33] Strauss, T. and S. Norman (2017): Modelling Chaoborus crystallinus populations to simulate effects and recovery under beta-cyfluthrin exposure. gaiac - Research Institute for Ecosystem Analysis and Assessment; Kackertstr. 10, 52072 Aachen, Germany. Report No. R-37809. Sponsor: ADAMA Makhteshim Ldt., Beer-Sheva 84100, Israel. Confidential Report.

among freshwater macroinvertebrates incl. oligochaeta, mollusca, hirudinea, cnidaria, nematocera, trichoptera, ephemeroptera and heteroptera. Further single-species laboratory testing revealed that larvae of *Chaoborus* sp. were the most sensitive organisms among members of amphipoda, trichoptera, megaloptera and plecoptera (Strauss and Norman 2017).

In these acute toxicity tests, L4 larvae were exposed to a series of nominal concentrations (0, 1.6, 5, 16, 50 or 160 ng/L) in test vessels with sediment, and acute mortality was observed after 96 h (Pickering 2016). Measurements of pesticide concentrations showed that in these vessels beta-cyfluthrin effectively dissipated from the water phase within 24 h. L1 larvae were exposed to 0, 1.6, 2.5, 5, 16, and 50 ng/L for 72 h in test vessels without sediment, but soaked alder leaves at its bottom (Cockroft 2017). Probably due to the absence of sediment, beta-cyfluthrin took slightly longer to dissipate from the water phase in these test vessels, and after 48 h, all larvae had died in all test concentrations (except of control).

The available experimental data cover most potentially relevant freshwater macroinvertebrate species in the simulated Central European ditches and ponds. The data clearly suggest *Chaoborus* sp. as most sensitive taxon, justifying the selection of *Chaoborus crystallinus* as model species despite of its potentially low vulnerability due short generation times (as compared to other freshwater insects).

3.6.4 The Environmental Scenario

The IBM *Chaoborus* population model was used to simulate effects of the application of beta-cyfluthrin to winter wheat in ponds and in edge-of-field ditches in the EU Central Zone. Simulations started on 1st January and were run for 1 year. Two pesticide applications were simulated on 15. May and 29. May. Based on the proposed GAP for beta-cyfluthrin, the modellers considered these dates broadly-representative for the EU Central Zone. The dates are located in the first half of the possible time window for applications. The modellers justified the selection of early dates with their expectation of maximum effects on *Chaoborus* populations in their main breeding season. However, later applications may result in more long-term effects because the time for population recovery before the end of the breeding season is then shorter. Therefore, an additional simulation with late application would have been useful as supporting information. In order to meet the desired ERO-RAC (see above), Strauss and Norman (2017) did not use PEC_{sw} values from FOCUS exposure modelling but lower PEC_{sw} values of 2.5 ng/L for ditches and 0.75 ng/L for ponds, assuming that these values can be achieved by further mitigation measures.

Ponds were considered to be isolated by distance. Conservatively, a 50 % loss of adult females due to emigration, but no immigration was assumed. The assumption that immigration can be neglected seems justified, considering that exchange between distant ponds is hardly predictable. A loss of 50 % of adult females seems very conservative and might be an overestimate, considering that females might return if they cannot find neighbouring ponds. The high migration rate of 50 % was assessed from a microcosm study where many artificial ponds were available in close vicinity (Strauss et al. 2016). In contrast, the ditch scenario considered 50 % migration of adult females between equally-distributed exposed and non-exposed ditches (1:1). The assumption of an equal number of exposed and non-exposed ditches seems reasonable worst-case because not all ditches in an area may be exposed to beta-cyfluthrin. However, the assumption that 50 % of emerged females migrate between exposed and non-exposed ditches may be too high, considering that in large-scale agricultural areas, midges may have to pass several hundred meters to reach a ditch neighbouring a different field.

The simulation of migration might have allowed to study indirect effects of pesticide application on non-exposed ditches that result from a reduced migration of individuals from exposed to non-exposed sites. However, the connected sites were not compared to an independent non-exposed site that serves as control. Instead, the connected non-exposed site was used as control scenario for the ex-

posed site. Therefore, simulated effects may have been underestimated, because the population directly affected from exposure was compared to an indirectly affected population instead of a non-affected population.

The only pesticide effect modelled was acute mortality. Sublethal effects on growth and reproduction – which are likely to occur at concentrations that are lethal for parts of the population and may undermine population recovery – were not considered. Additionally, information on delayed effects such as increased mortality during transitions from one life stage to the next seemed to be not available and thus not considered.

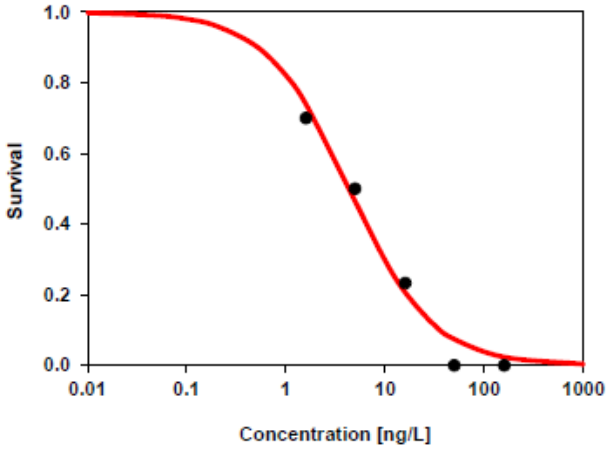
3.6.5 Parameter Estimation

The IBM *Chaoborus* population model was used with the default biological parameterization on the basis of data from laboratory standard tests and mesocosm studies (see the general model description in section 2.6). The individual-level effects of beta-cyfluthrin were implemented using a dose-response function instead of a TKTD approach which was justified by the modellers owing to the short DT_{50} of the pesticide in the water column (2.4 – 3.8 h) and the acute mode of action (Strauss and Norman 2017). These pesticide properties of beta-cyfluthrin may enable the use of a dose-response approach, assuming that effects observed in acute tests were driven by a very short exposure peak and do not increase over extended time periods. In these tests, concentration-dependent larval mortality was reported only by the end of a test after 72 or 96 h. However, the assumption that mortality will not increase after the end of the test was not tested, and individuals immobilised during the pulse exposure may have also recovered by the end of the test. These individuals would have not been considered affected, though they should be considered "ecologically dead" due to expected effects on their fitness (EFSA PPR 2018). Therefore, individual-level effects may have been underestimated and the use of a TKTD module such as GUTS might have been more appropriate to fit and predict effects after variable exposure times, but would require additional data.

For the daily mortality of L3 and L4 larvae in the population, a sigmoid dose-response function was fitted to the mortality observed after 96 h exposure in the acute test. This might result in simulated effects that are too high, because individuals in the model are four times subjected to a daily mortality rate that actually resembles cumulative mortality by the end of the test. However, in the population model individuals die when their individual tolerance value (randomly drawn between 0 and 1) is lower than the mortality from the dose-response function (Strauss 2017). Because this tolerance value is drawn only once at birth, individuals that survive exposure the first day will also survive exposure to the same or lower concentrations in the future. This implementation bears the risk of underestimating effects of exposure that is repeated or extended beyond the duration of the acute toxicity test used for parameterization. Extended exposure was no issue in this study due to the fast dissipation of beta-cyfluthrin. However, it was not established whether effects of the two exposure pulses are actually independent or if the sensitivity of *Chaoborus* larvae may increase following repeated exposure. The dose-response curve could be fitted well to the data, but the exposure of 0.75 ng/L considered in the pond scenario was lower than the lowest test concentrations (1.6 ng/L) used to establish the curve. Effects at 0.75 ng/L therefore had to be extrapolated which is associated with higher uncertainty than interpolation.

No dose-response curve could be fitted for L1 larvae, because all exposed individuals died in the toxicity test after 72 h. In the simulations, the modellers thus conservatively assumed that pulse exposure kills 100 % of the L1 / L2 larvae, irrespective of the considered concentration (2.5 or 0.75 ng/L). Therefore individual-level mortality on young larvae might have been overestimated in the pond scenario. A DT_{50} of 4 h was applied to model the dissipation of beta-cyfluthrin which matched the observations in the test of the L4 larvae and the cited DG SANCO Official Review Report.

Figure 35: Application of the IBM *Chaoborus* Population Model to beta-Cyflutrhin – Parameterization of Individual-Level Effects

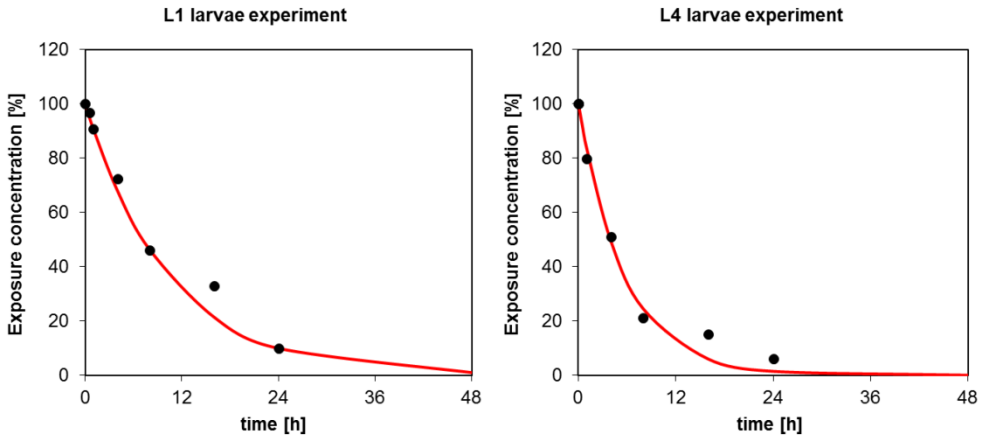


Dose-response curve for the survival of L4 larvae in test vessels with sediment after 96 h. Graph reproduced from Strauss and Norman (2017).

The environmental conditions, i. e. water temperature and food levels for young and older larvae, were adapted to meet conditions in the EU Central Zone. Water temperature was taken from a 15 year series of measurements in artificial ponds in Aachen which coincided with measurements from outdoor microcosms in Suffolk UK, but was about 14 days earlier in spring-warming and autumn-cooling than a standard lake water temperature profile for the 51° geographical latitude obtained from Straskraba & Gnau in 1983 (Strauss and Norman 2017).

At the beginning of the simulation in winter, only older larvae were considered to be present, which is typical for winter conditions. The initial densities of the L4 larvae were set to 1,000 – 3,000 individuals (1 to 3 larvae/L in ponds). Food saturation levels between 20% and 50% for all larval stages were chosen as a typical range derived from different mesocosm studies.

Figure 36: Application of the IBM *Chaoborus* Population Model to beta-Cyfluthrin – Dissipation of beta-Cyfluthrin in Acute Toxicity Tests



Dissipation of beta-cyfluthrin from the water phase in acute toxicity tests on young *Chaoborus crystallinus* larvae without sediment (left) and on older larvae (right) in the presence of sediment. Graphs reproduced from Strauss and Norman (2017).

3.6.6 Sensitivity and Uncertainty Analysis

It has not been tested which model parameters most strongly affect the regulatory relevant endpoint in this study, i. e the time for population recovery after pesticide application. Therefore, a full uncertainty analysis was not performed. However, for the ditch scenario simulations were run with and without immigration to assess the impact of this potentially highly relevant process on recovery time (see below).

For the regulatory model that has been applied to beta-cyfluthrin, no sensitivity analysis has been presented. However, a local sensitivity analysis for the model parameterized to the slightly different default environmental scenario without pesticide application was available (see the general model description in section 2.6 for details). Because parameterization of the population model was changed only slightly compared to the default scenario, the repetition of a sensitivity analysis would have probably not revealed much new information.

3.6.7 Comparison with Measurements

In the modelling report, predictions of the regulatory model for beta-cyfluthrin were not compared with observed results. However, model predictions on population recovery have been tested under a different exposure scenario with a different pesticide (see the general model description in section 2.6). It may have been interesting to adjust the environmental scenario to the microcosm studies of Heimbach (2000) and Jenkins (2014) in order to identify whether the model was able to reproduce the observed delay in emergence (data on larval abundance were scarce and probably not suitable for comparison). However, the data suggest that population recovery was driven by recolonization from neighbouring control microcosms. Conclusions from such a comparison on the performance of the model in the simulated field scenarios are therefore not straightforward, because recolonization heavily affected model predictions (see below) and the recolonization potential in the microcosm study was clearly different from the unknown and variable recolonization potential in the field.

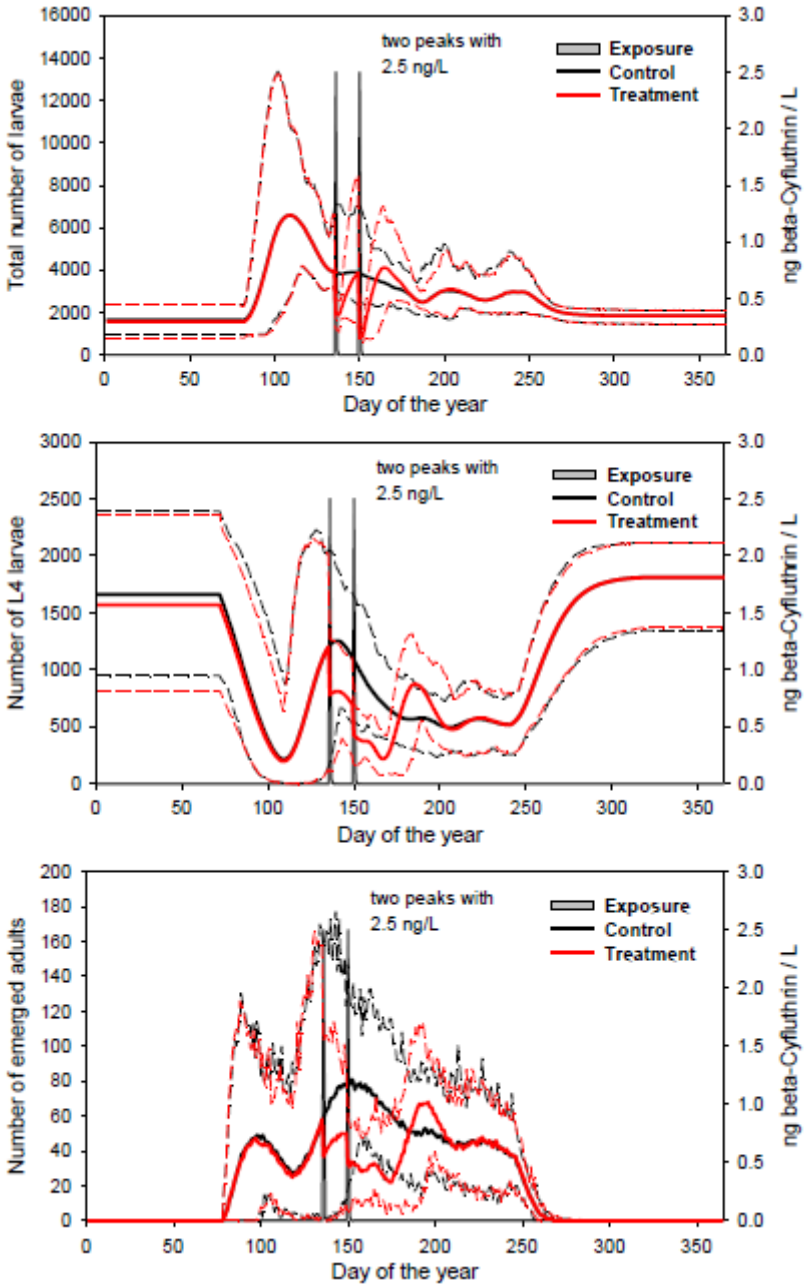
3.6.8 Model Use

The modelling report presented results on three endpoints: the overall sum of larvae, the number of older (L4) larvae, and the number of emerged adults over time. An endpoint was considered to recover when its average value (across 100 repeated simulations) in the exposed populations reached or exceeded the average value in the non-exposed populations.

The graphs indicate 3-4 peaks for the sum of larvae, the number of L4 larvae and of the emerged adults in control populations. This observation suggests that the simulations comprised 3 - 4 generations per year. This matches observations in the literature (Janz et al. 2016) and is quite common under Central Zone conditions according to the authors of the modelling study (personal communication), though in the modelling report a univoltine life-history has been reported (Strauss and Norman 2017).

In the ditch scenario, the L4 larvae were moderately reduced by beta-cyfluthrin on the days of application. Increased mortality of L1 and L2 larvae resulted in a further reduction in L4 larvae and emerging adults compared to controls in the following days. The overall sum of larvae recovered 24 d after the first pesticide application, the number of L4 larvae after 41 d and the number of emerged adults after 49 d. After reaching the control level, a temporary overshoot in the abundance of older larvae and adults occurred. Such a delayed numeric response to changes in the level of competition can be often observed in populations.

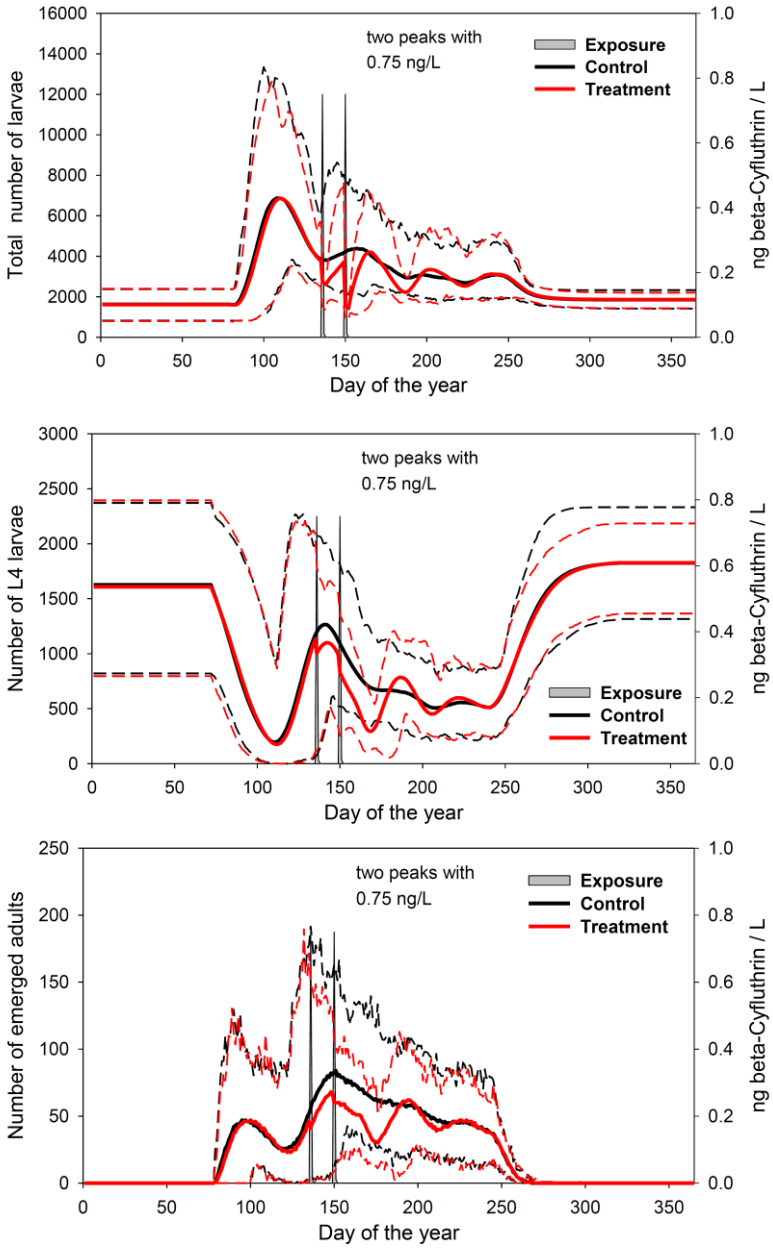
Figure 37: Application of the IBM *Chaoborus* Population Model to beta-Cyflutrhin – Simulations for the Ditch Scenario



Lines indicate the mean, minimum and maximum of 100 Monte Carlo simulations. Graphs reproduced from Strauss and Norman (2017).

In the pond scenario, the overall sum of larvae took slightly longer to recover (28 d) although the applied pesticide concentration was lower. The same was observed for the number of L4 larvae (44 d) and the number of emerged adults (54 d). Again, reaching the control level was followed by a temporary overshoot in the abundance of L4 larvae. The delayed population recovery in the isolated ponds can be explained by the lack of immigration of adults. This becomes apparent when considering the results for the additional isolated ditch scenario.

Figure 38: Application of the IBM *Chaoborus* population Model to beta-Cyfluthrin – Simulations for the Pond Scenario



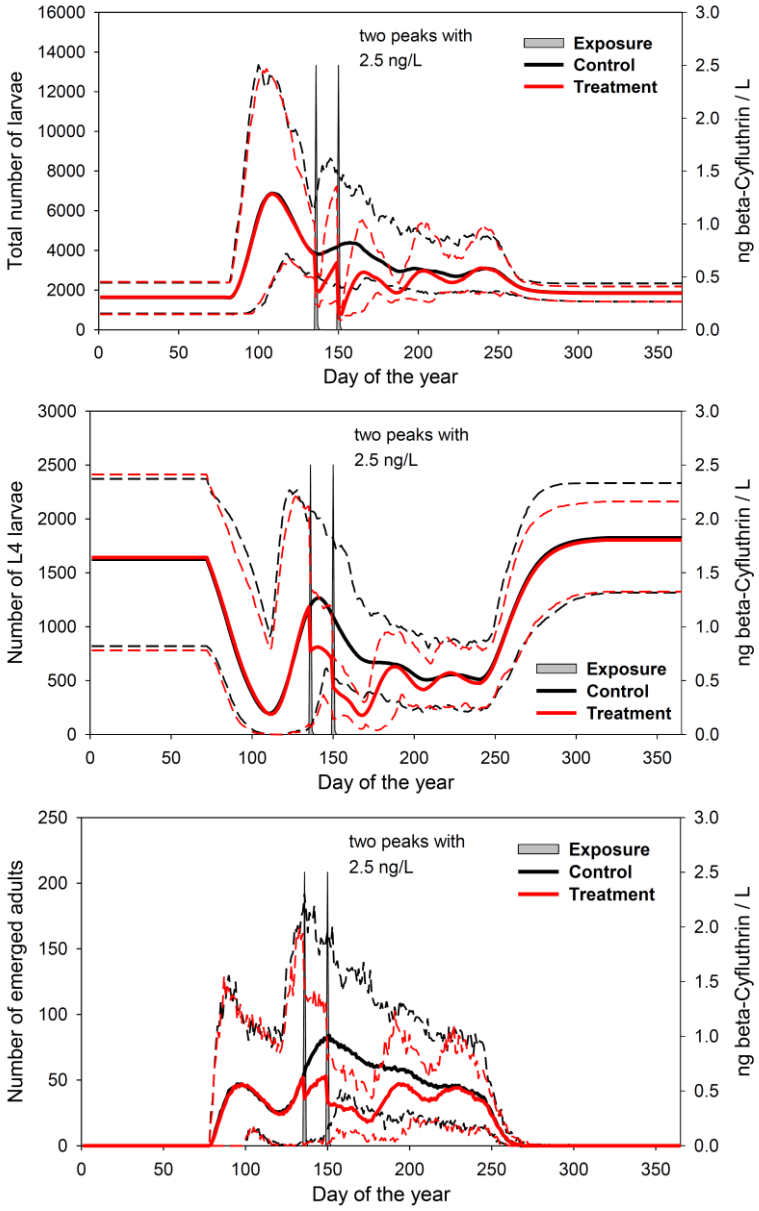
Lines indicate the mean, minimum and maximum of 100 Monte Carlo simulations. Graphs reproduced from Strauss and Norman (2017).

In the isolated ditch scenario, the number of L4 larvae took even a bit longer to recover (50 d) than in the pond scenario. As expected, the delay in recovery compared to the non-isolated ditch was particularly prominent for emerged adults (90 d) and the sum of larvae (70 d), but not so prominent for L4 larvae. This can be explained by the lack of immigrating adults that lowered the subsequent production of L1 larvae which had all died from the pesticide application.

In the modelling report of Strauss and Norman (2017), results were summarized in data tables that show the average time across repeated simulations until populations recovered to various percentage levels of deviation from control population size. This is a useful approach but may be supplemented with a measure of uncertainty such as 95 % confidence intervals. Overall, the results illustrate that in

the simulations, immigration was a main driver of population recovery: In the ditch scenario, immigration almost halved recovery time from 93 to 54 days after the first pesticide application. Because the assumed level of immigration is associated with high uncertainty, results from model runs with immigration should be considered with great care. It may have been useful to provide a graph of the modelled immigration rate [%] vs. time to full recovery to control population size [days, mean ± 95 % confidence]. This may have provided a margin of safety for the uncertainty in model predictions associated with the unknown immigration rate.

Figure 39: Application of the IBM *Chaoborus* Population Model to beta-Cyfluthrin – Simulations for the Isolated Ditch Scenario



Lines indicate mean (minimum and maximum) of 100 Monte Carlo simulation. Graphs reproduced from Strauss and Norman (2017).

3.6.9 Overall Judgement

From the modelling results, the authors concluded that recovery of all studied endpoints (abundance of young larvae, abundance of all larvae, and abundance emerged adults) takes 54 days in ditches with considerable recolonization from immigrating adults. Without recolonization, the recovery time extends to 93 days in ditches and to 86 days in less contaminated ponds. The EFSA guidance for aquatic risk assessment requires that populations have fully recovered within 8 weeks (56 d) after the first pesticide application (EFSA PPR 2013). Considering the high uncertainty associated with the assumed level of recolonization, the modelling results suggested a low margin of safety for the risk assessment.

The authors concluded that the predicted recovery times were very conservative due to the assumption of no immigration in the pond scenario and of 100 % acute mortality of L1/L2 larvae. Indeed, assuming no recolonization from immigration but at the same time 50 % loss of adults due to emigration is probably unrealistic worst-case. However, the high emigration rate affected both control and exposed simulations and thus may have underestimated the absolute population sizes, but not the difference between both scenarios. Given the isolation of ponds in agricultural landscapes, a low immigration rate is likely. Instead of only switching immigration on or off, an uncertainty analysis might have been helpful that evaluates the effect of varying immigration rates on the predicted recovery time. This way, conclusions from the modelling based on less extreme assumptions may have been informed.

The assumption of 100 % mortality of L1 larvae seems justified based on the acute toxicity tests of Cockroft (2017). All larvae died at the lowest test concentration of 1.6 ng/L which was higher than the concentration in the simulated ponds (0.75 ng/L). However, additional stressors may have increased the larval sensitivity in the field, so that the assumption of very high mortality also at 0.75 ng/L in the field seems realistic. Additional stressors may also increase the sensitivity of L4 larvae in the field so that effects on old larvae may have been underestimated.

The potential for an overestimation of effects is contrasted by other sources of uncertainty that lead to a potential underestimation of effects. These sources of uncertainty are related both to the general population model and to the specific model application. Uncertainty related to the general model includes the fact that no biotic or abiotic factors such as competitors and oxygen stress were considered that limit population growth and thus potentially delay population recovery. Additionally, in the model the elimination of L1 larvae from acute effects did not result in starvation of cannibalistic older larvae that prey on young larvae. See the evaluation of the general model in section 2.6 for further discussion.

Uncertainty related to the specific model application includes the fact that potential sublethal and delayed effects (e. g. on maturation time or emergence) were not incorporated which may significantly delay recovery and are likely to accompany acute effects. In particular, individual-level effects were fitted to mortality observed after 96 h when immobilized individuals may have recovered from acute immobilization but suffered likely from sublethal effects. Additionally, early pesticide application (compared to possible applications according to the GAP) was considered. Later applications may have resulted in recovery times extending to the next year due to less time for reproduction before the end of the breeding season. Finally, individual-level effects were implemented such that individuals that survived the first pulse exposure were also not affected by the second pulse of the same concentration, though effects may have built up after the second exposure.

Overall, uncertainty leading towards a potential underestimation of the real risk may have been outweighed by the uncertainty leading towards a potential overestimation of the real risk in this modelling study. Therefore, conclusions should be drawn with care, and considering the low margin of safety identified, the modelling results may not suggest an environmentally safe use. Nevertheless, despite the potential for underestimation of the real risk, the case study illustrates that the IBM *Chaoborus* population model is actually able to predict population effects over several months under field conditions.

3.7 MASTEP – Application to Pyridalyl

Evaluation by Jeremias Becker

3.7.1 Background

With the development of pyridalyl in 2002, a novel class of insecticides was introduced for the control of lepidopteran and thysanopteran pest species. Pyridalyl shows unique symptoms that suggest a distinct mode of action in comparison to previously registered insecticides (Sakamoto et al. 2004).

In 2006, Sumitomo Chemical Company Ltd. applied for the registration of pyridalyl as active substance in the European Union for use on cotton, fruiting vegetables and lettuce (Baveco et al. 2012). A microcosm study revealed high sensitivity of the water louse *Asellus aquaticus* to pyridalyl. Therefore Van den Brink et al. (2007b)[34] submitted a modelling report to the environmental agency of the Dutch EU Rapporteur Member State (Dutch Board for the Authorisation of Plant Protection Products and Biocides, CTGB) in which potential ecological effects of pyridalyl were investigated with the newly developed MASTEP population model. This report was one of the first Higher Tier studies using mechanistic effect modelling that have been submitted for governmental risk assessment of plant protection products in Europe. Some results of this study were also reported in the first scientific publication on MASTEP (Van den Brink et al. 2007a).

In 2011, the authors of the study added two annexes to the report. Annex 1 (Van den Brink et al. 2011b) provides correction of a programming error; all simulations have been repeated and compared to those in the original report. In annex 2, the authors presented additional simulations in which the input effects at organism-level were related to mid-term exposure in water only instead of short-term exposure in water and long-term exposure in sediment, following a request of CTGB (Van den Brink et al. 2011c). On request of EFSA, the authors finally submitted a revised report (Baveco et al. 2012)[35] incl. an updated FOCUS exposure modelling which served as input for MASTEP. The revised report also presented results of additional simulations using an environmental scenario without non-treated refuge areas, as compared to the standard scenario used in the report. Finally, the revised report included a sensitivity analysis on the impact of selected parameters on recovery time and an updated model description following the ODD standard (Grimm et al. 2006).

In its draft assessment report, CTGB (2012) concluded that in principle the MASTEP model is a scientifically sound approach, but that for this case study the available data were not sufficient, particularly regarding the exposure route of pyridalyl. In this review we focus on the modelling as described in the revised report of Baveco et al. (2012) that was used for the finally successful registration of pyridalyl.

3.7.2 Problem Definition

Pyridalyl has been characterized by fast adsorption but slow degradation and chronic re-sorption (CTGB 2012). Because pyridalyl failed Tier 1 risk assessment for aquatic invertebrates, a microcosm study (Springborn 2006) on sediment organisms (chironomidae, oligochaeta), gastropoda and *Asellus aquaticus* was performed (CTGB 2012). The established NOEC was higher than predicted surface water concentrations using FOCUS step 4 with mitigation measures (up to 0.05 µg/L) for all taxa except *A. aquaticus*. In contrast, effects on *A. aquaticus* were observed even at the lowest test concentration so

[34] Van den Brink, P. J., I. Roessink, P. I. Adriaanse, W. H. J. Beltman, J. Verboom, P. Schippers and H. Baveco (2007): Landscape Simulation of Temporal and Spatial Effects on Asellus aquaticus from Concentrations of Pyridalyl in Surface Water. Alterra Wageningen University and Research Centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0058. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome, Chuo-ku, Tokyo 104-8260, Japan. Confidential Report.

[35] Baveco, J. M., W. H. J. Beltman, G. Fait and P. J. Van den Brink (2012): Estimating recovery times of Asellus aquaticus after pyridalyl exposure by the MASTEP model. Alterra, Wageningen University and Research centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0123. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome,Chuo-ku, Tokyo 104-8260, Japan.

that no NOEC for this species could be established and consequently no RAC could be derived from the microcosm study.

On day 72, new individuals of *A. aquaticus* were introduced to the microcosms to simulate recolonization. However, introduced organisms did not clearly establish, suggesting that in addition to an acute population decline, chronic effects existed from the pesticide that was still present in the system (mainly bound to the sediment). However, the introduction of new individuals may not represent realistic rates of recolonization. Therefore, the applicant submitted a simulation study using MASTEP to provide a line of evidence that pyridalyl causes no unacceptable long-term effects on populations of *A. aquaticus* in real-world freshwaters which are subject to recolonization. This way, *A. aquaticus* might be excluded from the NOEC calculation based on the microcosm study so that pyridalyl can pass Tier 3.

3.7.3 Supporting Data

The basic population model of MASTEP was developed using expert judgement and experimental data on life history and movement of *A. aquaticus* (see evaluation of the general model in section 2.5). Exposure and effects of pyridalyl were parameterized using the microcosm study of Springborn (2006) wo tested effects after 4 applications of pyridalyl (apparently with a 7 d interval) with the following nominal concentrations in water: 0, 0.05, 0.1, 0.65 and 6.5 µg/L. Exposure to 0.65 µg/L and higher resulted in an acute population decline in *A. aquaticus* one day after the first application; exposure to 0.05 µg/L and higher caused a chronic population decline 28 - 56 days after the first pyridalyl treatment (i.e., without recovery until the end of the study (98 d). At day 72, 20 % of the population size in the control was re-introduced to the three highest treatments. In contrast to pyridalyl concentrations in water, concentrations in sediment did not considerably decrease over the course of the study.

According to the CTGB, the microcosm study generally fulfilled scientific standards. However, the study did not provide sufficient information on the exposure routes leading to the observed chronic effects. Additional tests in 2008 suggested that chronic effects on *A. aquaticus* were not related to chronic exposure in sediment, but to chronic exposure in water. This motivated the CTGB to request that chronic effects in MASTEP should be linked to chronic exposure in water instead of sediment (see Van den Brink et al. 2011c).

The CTGB (2012) considered the microcosm study to represent a realistic freshwater community. However, the study did not cover various important insect taxa in streams and ditches such as ephemeroptera and plecoptera which are known to be particularly vulnerable to pesticides (Liess and von der Ohe 2005). Therefore, conclusions based on modelling with *A. aquaticus* as the most sensitive taxon in the microcosms may not be protective for the full macroinvertebrate community in real streams and ditches which can include additional species that are similarly sensitive but more vulnerable.

3.7.4 The Environmental Scenario

Simulations in MASTEP were run with a spatial resolution of 1 m^2 and a temporal resolution of 1 d. Considering the intended application of pyridalyl, effects on *A. aquaticus* were simulated in ditches and streams (Baveco et al. 2012). According to the authors, ponds were not simulated because pond scenarios did not exist for Southern Europe and because applications in Northern Europe were restricted to protected crops in greenhouses only (Van den Brink et al. 2007b).

Ditches and streams were modelled as an exposed stretch of 1 m width and 100 m length, followed by a none-treated refuge stretch of 400 m length downstream. Both ends of the water body were connected so that individuals could “jump” from one end to the other. The simulations thus represented a random section of a water body that is composed of 100 m exposed stretches neighbouring a treated agricultural field that are separated by 400 m non-exposed stretches. This way, simulations covered allogenic population recovery in exposed stretches via recolonization from refuge stretches that are

not directly exposed; additionally, the simulations considered indirect exposure of the refuges by downstream drift of pyridalyl, which may affect recolonization. The landscape composition may not be conservative for intensely used areas in which the length of agricultural fields can exceed 100 m. This was demonstrated when the standard scenarios were compared with modified landscape scenarios that covered only the 100 m exposed stretch. Results showed that refuges had a significant effect on population recovery in streams, as long as the drifting of animals with the current was simulated (Baveco et al. 2012).

Following the critical GAP, four applications of 150 g pyridalyl/ha with 7 d intervals were considered. The daily pyridalyl concentration in each grid cell of the MASTEP landscape was calculated from external FOCUS fate modelling prior to the simulations, which was generally well justified. In a first step, various FOCUS step 3 simulations were run to identify the worst-case scenarios to be used for MASTEP modelling: The predicted time weighted average exposure concentrations of 28 days (TWA28d) in water was calculated for different stream and ditch scenarios in the EU Southern and Northern zone using TOXSWA. No FOCUS guidance was available for the application in greenhouses in Northern Europe. The authors therefore followed a request of the CTGB considering 0.1 % spray drift while discarding drainage and runoff loadings. Additionally, in Northern Europe FOCUS scenarios for leafy vegetables were used (modified with application rates for fruity vegetables), because no FOCUS scenarios for fruity vegetables had been defined. For Southern Europe, highest TWA28d were identified for the application in cotton in the D6 ditch scenario and for the application in fruiting vegetables in the R3 stream scenario. In Northern Europe, highest exposure was identified for the application in fruity vegetables in the D3 ditch scenario.

In a second step, various FOCUS step 3 and step 4 exposure profiles were calculated for these three worst case scenarios to predict daily concentrations (TWA1d) that were afterwards used as input for MASTEP. For the D6 ditch, step 4 calculations considered a 10 m buffer zone to reduce spray drift. For the R3 stream, three alternative mitigation measures were considered: 10 m vegetative buffer strip leading to reduction in spray drift as above plus reduction in run-off (60% reduction in aqueous, 85% reduction in erosion); 20 m vegetative buffer strip leading to reduction in spray drift as above plus higher reduction in run-off (80% reduction in aqueous, 95% reduction in erosion); 10 m buffer zone to reduce spray drift as above plus 100 % reduction in run-off. For the D3 ditch, exposure was reduced as described above due to the use in greenhouses.

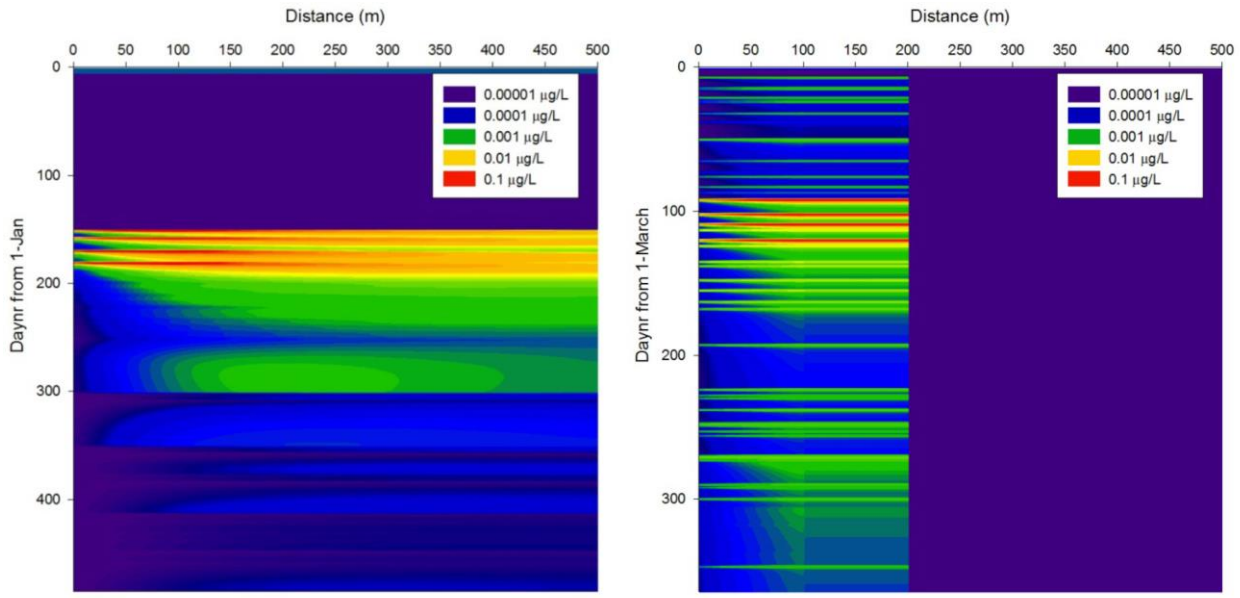
3.7.5 Parameter Estimation

MASTEP was applied with the default parameterization of the basic population model as described in Van den Brink and Baveco (2009). Daily pyridalyl concentrations in water (TWA1d) for each landscape cell were calculated prior to the MASTEP simulations using an extended TOXSWA model. In the revised modelling report of Baveco et al. (2012), parameterization of pesticide properties and the time window for pyridalyl application in TOXSWA followed the updated EFSA recommendations (see above).

The extended TOWSXA model also considered exposure via downstream drift from directly contaminated cells. For the ditch scenarios, drainage water was assumed to enter laterally across the overall simulated 500 m stretch, but contained pyridalyl only in the first 100 m. In contrast, the stream scenarios could not be extended without changing the hydrologic parameterization. Instead, the authors calculated exposure in three sections of the modelled stream stretch separately, using different methods: Exposure in the first 100 m followed the normal FOCUS modelling. Exposure in the 100 – 200 m section was fixed to a value obtained from a separate FOCUS simulation. This simulation provided the concentration at the end of a 100 m stretch in which only the first 90 m are directly contaminated. Concentrations in the 200 - 500 m section were taken zero, assuming that pyridalyl had been fully adsorbed to the sediment in the first 200 m (Van den Brink et al. 2007b). This scenario is worst case for the 100 – 200 m section, but best case for the 200 – 500 m section. However, predicted exposure for

the ditch scenario that is more in line with the standard FOCUS methodology shows that pyridalyl concentrations do not considerably decrease in the 100 – 200 m section and remain high for the entire modelled 500 m stretch (Fig. 40). Transportation to downstream sections may be even stronger in streams due to the higher flow velocity. Therefore, the exposure scenario in streams may have resulted in an underestimation of population effects, because best case assumptions for the longest part of the simulated stream stretch seem to outweigh worst-case assumptions for the 100 – 200 m section. Moreover, the best case 200 – 500 m section is more relevant for recolonization than the worst case 100 – 200 m section, because in the simulations, individuals enter the contaminated section mainly via downstream drift from the last 200 – 500 m.

Figure 40: Application of MASTEP to Pyridalyl – Predicted Spatiotemporal Exposure



Predicted daily pyridalyl concentrations in water (dissolved plus adsorbed to suspended solids) for the D6 ditch scenario (left) and for the R3 stream scenario (right) without mitigation measures (Baveco et al. 2012). The scenarios considered direct pyridalyl exposure in the 0 – 100 m section due to 4 applications between day 150 and 190. Later increases in exposure were related to a combination of remainders of the spray drift, drainage events and desorption from sediment in the ditch, and to run-off in the stream. Sharp decreases in pyridalyl concentrations in the ditch may have resulted from flushing events, but were not explained by the authors. In the stream, no pyridalyl dissipation was assumed for the 100 – 200 m section, and no exposure was assumed for the 200 – 500 m section. Graphs reproduced from Baveco et al. (2012).

Individual-level effects in MASTEP were implemented based on the results from the microcosm study of Springborn (2006). The observed population decline in the microcosms at a given observation time was directly related to the observed time-weighted average (TWA) concentrations using log-logistic regression. This dose-response function provided an endpoint that was called “ecotoxicologically relevant concentration” (ERC). Initially, the modellers assumed that the observed acute decline in population size resulted from peak exposure in the lowest 20 cm of the water phase, whereas the chronic decline resulted from long-term exposure in sediment (Van den Brink et al. 2007b). Accordingly, separate dose-response functions for acute and for chronic effects (ERCacute and ERCchronic) were established.

For acute effects, the observed population decline six days after the first peak exposure (first post-exposure sampling in the experiment) was related to the TWA1d concentrations in water at the time of

the first peak exposure (mean of concentrations measured or estimated 1 h and 1 d after contamination). Thus, it was assumed that effects occurred mainly at the first day due to the fast mode of action of pyridalyl. In MASTEP, the obtained ERCacute function was used to provide a daily mortality rate (additional to the background mortality) to each exposed individual based on the concentration experienced at the current (daily) time step. In case of multiple exposure peaks or extended exposure in the simulations, the acute effect was executed every day based on the currently experienced TWA1d, independent from previous exposure. This approach does not consider potential acclimatization (build-up of tolerance) and was considered worst-case by the authors. On the other hand, no potential build-up of effects within individuals with repeated or extended exposure was considered. Whereas the highest concentration (6.5 µg/L) immediately decreased the population size to < 1 % after the first treatment, the second highest concentration (0.65 µg/L) decreased population size only to ca. 50% after the first treatment, but to ca. 1 % after the second treatment (Fig. 42, left panel). This pattern suggests that a concentration-dependent build-up of effects may have occurred.

For chronic effects, survival 56 days after the first peak exposure (when maximum effect was observed) was related to the time weighted average concentration of the 56 days following the first exposure peak (TWA56d) in sediment (ERCchronic, corrected for organic matter). Only data from the control microcosms and the two lowest test concentrations were used in this case, because those showed no acute mortality. To calculate TWA concentrations for this dose-response function, first the concentration every hour was estimated from interpolation between the pesticide measurements in the microcosms, assuming degradation to follow first-order kinetics (Van den Brink et al. 2011c). Then the area under the curve of concentration vs. time was calculated as TWA. Thus, the dose-response function was parameterised using hourly TWA concentrations from the microcosms, but was applied in MASTEP to daily TWA concentrations provided by TOXSWA . This approach seems conservative because for the first day after pesticide application, the area under the curve is somewhat lower for the hourly compared to the daily TWA concentrations, where the same initial concentration starts to decrease later in time. Relating a given effect to a lower area under the curve will lead to a somewhat larger effect in MASTEP with a larger area under the curve for the first day, based on daily TWA calculations.

However, later tests suggested that chronic effects were not driven from exposure in sediment. Therefore, in the revised versions of the modelling report from Van den Brink et al. (2011c) and Baveco et al. (2012), both chronic and acute effects were linked to exposure in water. Because concentrations in water decreased rapidly after the last peak exposure in the microcosms at day 21, the authors assumed that the observed chronic population decline was actually caused by latent effects after exposure rather than by extended exposure.. The new modelling approach used a single dose-response curve for the overall observed population decline without differentiating between acute and chronic effects (general ERC). In this approach, survival 56 days after the first peak exposure was related to the time weighted average concentration of 28 days following the first peak exposure (TWA28), i. e. until one week after the last exposure peak. Thus, the general ERC related exposure that occurred mainly *during* the (four week) application window to effects observed four weeks *after* the last application. This approach might have underestimated effects due to a potential population recovery through reproduction after the last application. However, Fig. 42 (left panel) shows that populations in the microcosms exposed to the two highest concentrations did not recover from acute population decline before day 56. Additionally, populations exposed to the two lowest test concentrations started to decline not before the last application at day 21 and continued to do so at least until day 70, after which new individuals had been introduced to the microcosms. At the lowest test concentration, no individuals had been introduced and the populations did not recover until the end of the experiment. Therefore, recovery from acute population decline seems not to have affected the estimation of the full population effect after 56 days.

However, the ERCchronic and general ERC related TWA56 or TWA28 to the cumulative mortality after 56 days, while MASTEP required a daily mortality rate. Therefore, in the first report of Van den Brink et al. (2007b), the cumulative mortality was converted to daily mortality using the 56th root, assuming constant mortality over time. This way, pesticide-induced chronic mortality in the simulations slowly increased when exposure set in and dropped when exposure ended until it reached zero 56 days after last exposure (when TWA56 reached zero). In contrast, in Baveco et al. (2012), the conversion was not clearly described; probably the cumulative mortality at day 56 was translated to daily mortality across the 28 days considered for TWA28d by calculating the 28th root. In this case, the overall pesticide-induced mortality in MASTEP set in more slowly than the observed acute mortality and could not extend beyond a maximum of 28 days after exposure (when TWA28d becomes 0). This contrasts observations in the microcosms with lowest test concentrations where populations continued to decline until the end of the experiment (100 days after first exposure), long after exposure had declined to very low concentrations. If chronic effects were indeed driven by exposure during the application window, the duration of effects was simulated too short (latency of effects underrepresented). Additional data on the duration until the onset of population recovery would have been required to appropriately implement the latency of effects. In contrast, if chronic effects were driven by continuous exposure to low concentrations from re-sorption also after the application window, a different dose-response curve with such low concentrations would have been required.

From the available data, no distinction between both possible pathways can be made. In both cases, however, the implementation of effects may have resulted in an overestimation of the population recovery potential because the duration of effects was too short. This data gap should have been discussed more thoroughly by the authors. Nevertheless, simulated population effects with the new TWA28d approach were in most cases slightly more pronounced than those based on acute water and chronic sediment exposure (Van den Brink et al. 2011c). The reason is probably the fact that in the first approach based on ERCacute and ERCchronic, most of the pesticide-induced mortality (the acute mortality) acted immediately after exposure, giving time to population recovery afterwards. In contrast, in the second approach, the peak of mortality occurred later (when TWA28d reached maximum), so that also population recovery set in later.

Finally, the observed population decline in the microcosms was modelled exclusively as increased mortality in MASTEP. However, other important life history traits such as reproduction or the ability to migrate may have been affected as well. Additionally, it was not assessed whether life stages may have differed in sensitivity to pyridalyl. E. g., the combination of acute mortality on young individuals and chronic effects on the reproduction of mature individuals may severely affect the potential of a population to recover.

3.7.6 Sensitivity and Uncertainty Analysis

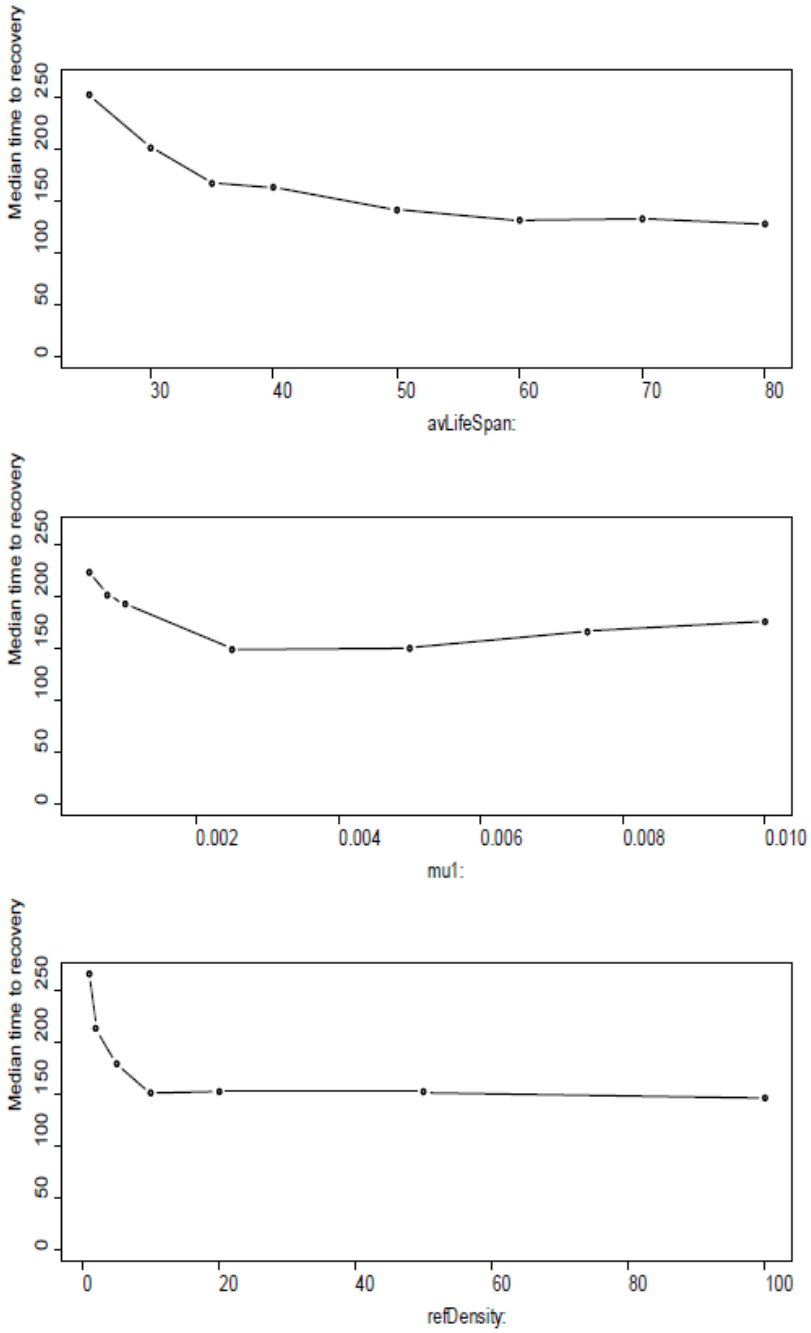
A local sensitivity analysis was submitted in the revised report of Baveco et al. (2012) in which one parameter at a time was varied. This way, the magnitude of the independent effects of each parameter on the studied endpoint could be investigated, while interactive effects of several model parameters remained unknown. For a non-exposed control run the impact of the changed parameter value on population equilibrium size was studied. For an exposure scenario, the impact on the time to recovery was studied, which is probably the most relevant endpoint from a risk assessor's perspective. According to the authors, a simple hypothetical exposure scenario was used to achieve sufficient generality in the results.

Three parameters were analysed that are notoriously unknown or associated with high uncertainty because they relate to background mortality and to the strength of density-dependence: the average lifespan (reciprocal of the daily background mortality risk), the coefficient of density-dependent mortality, and the reference density used to calculate the density-dependent clutch size of an *Asellus* brood. To avoid the confounding influence of seasonality, drift and migration from refuges, a simplified

South European ditch scenario was used for the sensitivity analysis with a stream length of 100 m and a single homogeneous exposure event (90 % mortality everywhere).

The analysis showed that only very small values of all three parameters below those values used in the regulatory model had a large effect on both studied endpoints. However, the presentation of the outcomes should be improved by showing the parameter values of the regulatory model in the figures for reference. Additionally, it would have been interesting to extend the analysis to more parameters such as those related to the movement pattern of individuals which are also associated with high uncertainty.

Figure 41: Application of MASTEP to Pyridalyl – Sensitivity Analysis for Population Recovery



Sensitivity analysis for the effects of the average life span (reciprocal of the daily background mortality risk, top), the coefficient of density-dependent mortality (middle), and the reference density for the calculation of the density-de-

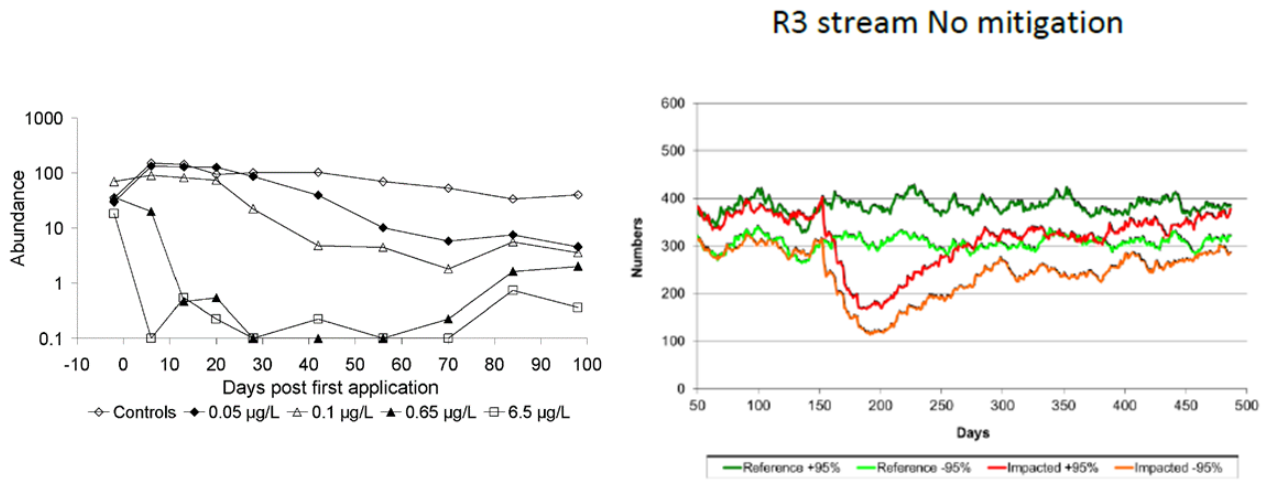
pendent clutch size (bottom) on the recovery time after a spatially homogeneous 90% mortality event. Parameter values used in the regulatory model: average life span = 40; density-dependent mortality rate = 0.005; reference density = 10. Graphs reproduced from Baveco et al. (2012)

3.7.7 Comparison with Measurements

When the modelling study was first proposed in 2007 and finalized in 2012, simulations of MASTEP in general had never been tested with real-world observations. Additionally, the toxicity module for individual-level effects makes assumptions on the reciprocity of concentration and exposure time when predicting effects from TWA concentrations. The same TWA concentrations can result from different exposure profiles experienced by each individual in the model; it has not been tested whether predictions hold for exposure profiles other than the one used for calibration (Miller et al. 2000). The study was submitted before the EFSA Sci. Op. on GMP (EFSA PPR 2014b) had been published which explicitly requires effect models to be validated with real-world observations. Nevertheless, considering the high uncertainties in model predictions outlined in the general model evaluation of MASTEP (see above), the application of an untested model for the governmental risk assessment of plant protection products may be considered premature.

Given that no validation with independent data was available for the model in general, predictions of the regulatory model may have been compared to the observed effects in the microcosm study used for parameterization. In the final version of the modelling study (Baveco et al. 2012), additional simulations were run that contained only the 100 m contaminated stretch. These scenarios may be compared to the microcosms which also provide no non-contaminated downstream stretch. The non-protected R3 stream scenario included four exposure peaks with ca. 0.1 µg/L (Fig. 40, right) that is in the range of the lowest test concentration in the microcosms. However, while this exposure resulted in ca. 90 % population decline after 98 d in the microcosm study, MASTEP predicted only ca. 30 % population decline 98 d after the first application (Fig. 42). Moreover, the trend in the microcosm study suggests that effects will remain constant or even increase after the end of the experiment, while MASTEP simulations suggested that populations were recovering after 98 days. The comparison suggests that the presented MASTEP simulations underestimated the chronic effects of pyridalyl.

Figure 42: Application of MASTEP to Pyridalyl – Validation with Microcosm Observations



Comparison of MASTEP model predictions with microcosm observations. Left panel: Effects of pyridalyl on the population size of *Asellus aquaticus* in the microcosm study of Springborn (2006), as presented in the first modelling report (Van den Brink et al. 2007b). Pyridalyl was applied at days 0, 7, 14 and 21. Note the logarithmic scale of the vertical axis. Graph reproduced from Van den Brink et al. (2007b). Right panel: Effects predicted by MASTEP for an R3 stream scenario that only consists of the 100 m stretch with comparable peak exposure (four pulses of ca. 0.1 µg/L with time intervals of ca. 1 week), as presented in the final modelling report. The upper and lower boundary of 95 % confidence

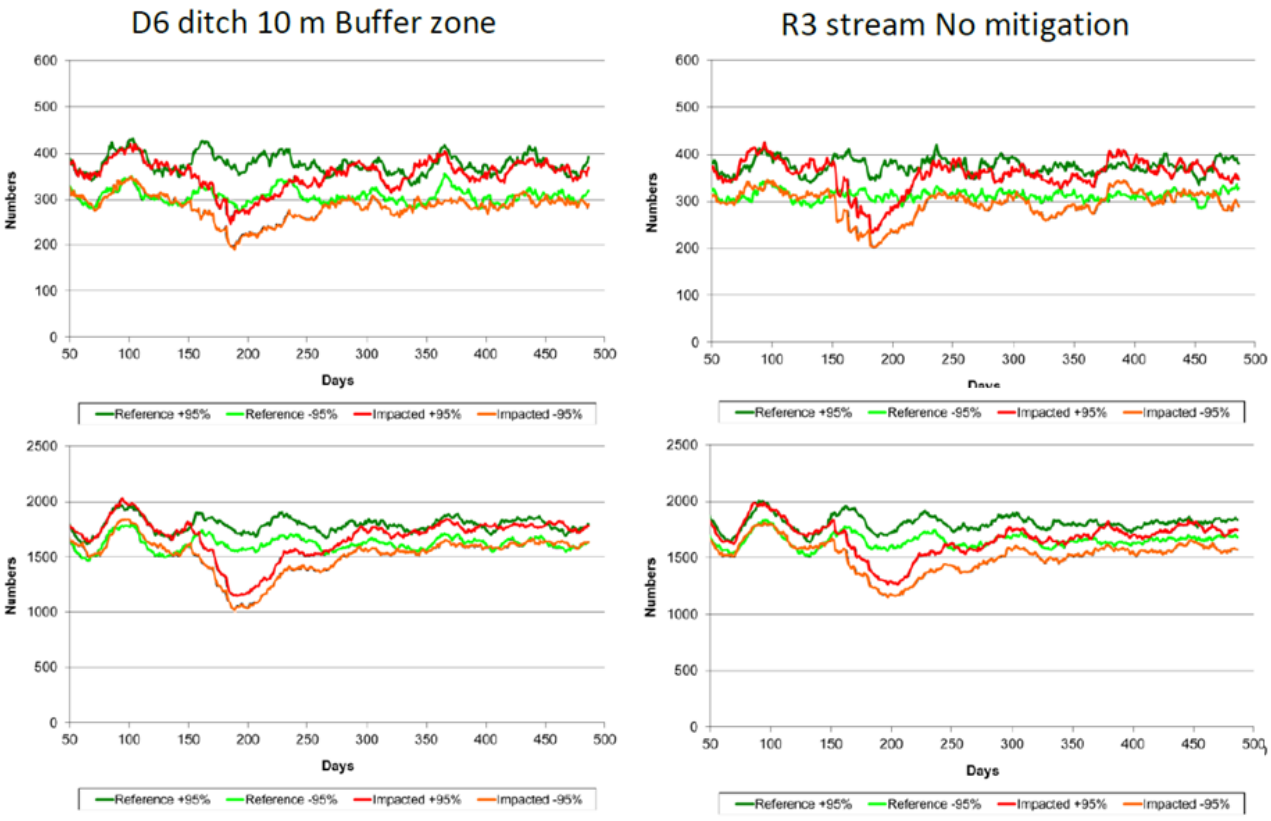
intervals for the total number of individuals from at least 20 replicate model runs are shown. Graph reproduced from Baveco et al. (2012).

3.7.8 Model Use

In the D6 ditch scenario without mitigation, acute effects immediately eliminated the population in the contaminated stretch in the first report of Van den Brink et al. (2007b); recovery from the non-contaminated stretch took ca. 6 months. A 10 m buffer zone decreased effects in the exposed stretch to a population decline over 4 weeks; no clear effects were observed for the overall stretch. Correction of the exposure equation in Van den Brink et al. (2011b) did not considerably affect model predictions. When chronic effects were linked to exposure in water, effects in the overall stretch increased in the D6 scenario with and without mitigation, while effects in the contaminated stretch decreased in the presence of a 10 m buffer strip (Van den Brink et al. 2011c). The update of the FOCUS exposure parameterization (Baveco et al. 2012) increased the population decline, but not recovery time in the overall stretch (Fig. 43).

In the R2 stream scenarios used in the first modelling report (Van den Brink et al. 2007b), a clear population decline was observed that lasted until the end of the study (150 d after first exposure), except when run-off was reduced by 100 %. Correction of the exposure equation in Van den Brink et al. (2011b) considerably decreased the predicted effects such that they remained only visible without mitigation measures. Linking chronic effects to exposure in water decreased effects further (Van den Brink et al. 2011c). No results were presented for the R2 scenarios after revision of the exposure parameterization in the final report, because then only the R3 stream scenario was considered worst-case.

Figure 43: Application of MASTEP to Pyridalyl – Simulations for Southern Ditch and Stream



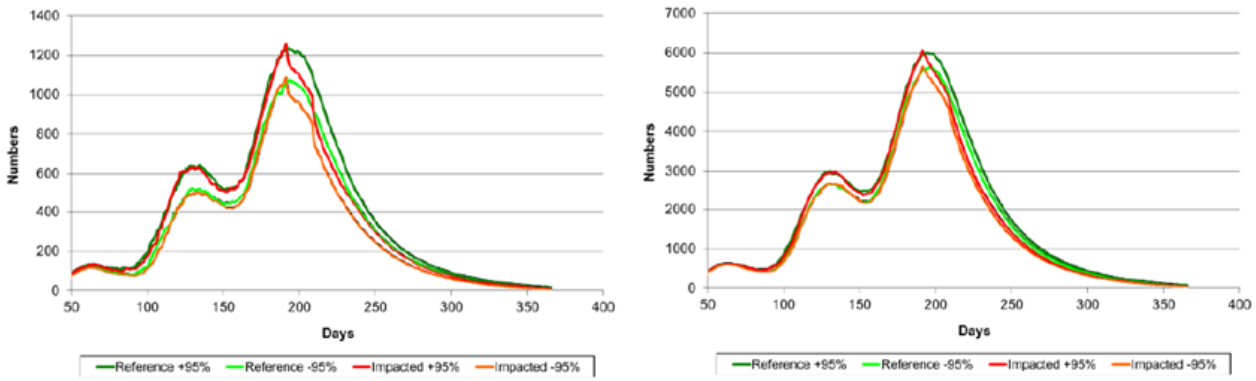
Simulated effects of pyridalyl in the southern ditch and stream scenarios. The graphs show the upper and lower boundary of 95 % confidence intervals for the predicted population size in the contaminated 100 m stretch (top) and

in the overall 500 m stretch (bottom) from at least 20 replicate model runs. Left panels show results for the D6 southern ditch scenario with 10 m buffer zone; right panels show results for the R3 stream scenario without mitigation. Graphs reproduced from the final modelling report (Baveco et al. 2012).

In the R3 stream scenarios, clear and persistent population decline until the end of the modelling study was observed only without mitigation measures in the contaminated and in the overall stretch in the first report (Van den Brink et al. 2007b). After correction of the exposure equation in Van den Brink et al. (2011b), the initial population decline in the same scenario was followed by recovery within 30 d in the contaminated stretch, and within 100 d in the overall stretch. Linking chronic effects to exposure in water decreased the magnitude of the acute population decline in the contaminated and the overall stretch, but the effect lasted longer (for ca. 50 and 150 d, respectively) and was even observed when run-off was reduced by 97.5 % (Van den Brink et al. 2011c). Updating the exposure parameterization did not change the outcome significantly (Fig. 43).

In contrast to the previous scenarios, population size in reference populations of the D3 ditch showed a distinct seasonal variation with two peaks in May and July (Fig. 44). No considerable effects of pyridalyl exposure were predicted in any of the submitted reports.

Figure 44: Application of MASTEP to Pyridalyl – Simulations for Northern Ditch



Simulated effects of pyridalyl in the northern ditch scenario. Predicted population size in the contaminated 100 m stretch (top) and in the overall 500 m stretch (bottom) of the D3 Northern ditch scenario without mitigation measures, as presented in the final modelling report. Graphs reproduced from Baveco et al. (2012).

The description of MASTEP and its application in the reports was generally well structured. In Baveco et al. (2012), a revised model description according to the ODD standard was presented. In Van den Brink et al. (2007b), the authors of the report suggested the following improvements: 1.) Laboratory studies to reduce the uncertainty concerning the chronic dose-effect relationship, which was considered the greatest improvement in the modelling because the exposure route of chronic effects (water, organic matter in sediment or macrophytes) is not known. 2.) A (semi-field) study with a closed mass balance accompanied by targeted laboratory experiments to quantify the main process parameters. 3.) More research to establish more realistic estimates of *Asellus* movement rates (presently based on one experiment which was performed in an artificial environment without food and shelter).

3.7.9 Overall Judgement

The modelling study was generally well documented. However, MASTEP was applied to a pesticide that exerted chronic effects, and thus used beyond the originally intended domain of applicability (fast-acting, non-persistent insecticides, Van den Brink et al. 2007a). General conclusions should be therefore drawn with care.

Additionally, results from the study are specific for the model species *A. aquaticus*. *A. aquaticus* was identified as the most sensitive species in the microcosm study of Springborn (2006) and therefore selected as model species. However, to our information this microcosm study did not include taxa such as ephemeroptera and trichoptera that are ecologically more vulnerable due to their longer generation times and typically also more sensitive to insecticides than the studied gastropoda, oligochaeta, chironomidae and *A. aquaticus*. Conclusions from the modelling study for the risk assessment of pyridalyl may be therefore not protective for various freshwater insect taxa.

The authors concluded from the modelling that chronic effects were much more severe for the population size over time than acute effects. This was not surprising, because acute mortality may be compensated by an increase in population recovery due to the release from intraspecific competition. In contrast, chronic effects from high short-term or from low long-term exposure may not release individuals from competitive stress but lower reproduction and decrease interspecific competitive strength as compared to less sensitive species. With additional information from chronic laboratory tests, it might have been more adequate to model such chronic effects as a pesticide-induced decrease in reproduction or other performance parameters, than as increased mortality.

In contrast to the level of detail spent on recolonization, various aspects that may considerably decrease population recovery have not been considered, e. g. species interactions, additional stressors and their interaction with individual-level pesticide effects. Additionally, individual-level pesticide effects in MASTEP were implemented in a way that linked chronic effects to acute exposure but limited effects to last only up to 28 days after exposure. This model behaviour seems inconsistent and was not supported by experimental observations in the microcosm study. Taken together, these issues raised the risk of a potentially unbalanced risk assessment. Interestingly, most of these conceptual issues have not been criticised by the CTGB during the process of risk assessment, in contrast to various more technical issues.

Finally, the model predictions were associated with considerable uncertainties that result from the unknown exposure route for chronic effects, from the potentially optimistic landscape composition and exposure modelling in the stream scenarios, and from limited knowledge on various aspects such as migration patterns in natural populations. These uncertainties could not be quantified in the study. In the modelling reports, simulated effects and recovery were not compared to the experimental observations in the microcosm study of Springborn (2006); however, such a comparison was done in the present evaluation and suggested that the risk of pyridalyl may have been underestimated in the modelling study.

When the modelling study was proposed for governmental risk assessment, MASTEP had been just developed and presented to the scientific community without time being spent in model validation and uncertainty analyses. This very early application of MASTEP was criticised in a review of Jarvis & Wynes (CTGB 2012) and suggests that additional scientific work may improve applications of MASTEP for the risk assessment of plant protection products.

3.8 MASTEP – Application to Deltamethrin

Evaluation by Jeremias Becker

3.8.1 Background

Deltamethrin is a type II pyrethroid insecticide that has been registered as plant protection product in the European Union for more than 10 years (EU Pesticides Database 2018). In an assessment report from the German Federal Environment Agency (UBA 2010), high risk was identified for several aquatic macroinvertebrates. Two Higher Tier studies from Schanné and van der Kolk (2001) and Heimbach et al. (2005, not available for this evaluation) addressed the effects of deltamethrin exposure from spray drift on freshwater communities in mesocosms. The studies showed significant effects of deltamethrin, formulated as emulsifiable concentrate, on glassworms (*Chaoborus sp.* larvae), ephemeropterans and on the water louse *Asellus aquaticus* even at low concentrations (10 ng active substance per litre). Considering not the nominal, but realistic concentrations observed or estimated at the ground of the mesocosms, *A. aquaticus* populations strongly decreased immediately after first pulse exposure to 10 ng/L, without proven recovery until the end of the studies (UBA 2010). High risk was expected also for other sensitive taxa such as *Gammarus sp.* that have not been tested in the mesocosm studies. Specifically, *Gammarus fasciatus* was identified as the most sensitive species in laboratory studies on deltamethrin (UBA 2010).

In order to demonstrate that affected populations can recover, the applicant Bayer CropScience AG presented a modelling study of Verboom et al. (2005)[36]. In this study, the Metapopulation model for Assessing Spatial and Temporal Effects of Pesticides – MASTEP (Van den Brink et al. 2007a) was applied to simulate autogenic (reproduction) and allogenic (recolonization) population recovery of the model species *A. aquaticus* after a series of deltamethrin applications.

3.8.2 Problem Definition

The aim of the modelling study was to relieve the mesocosm results by demonstrating that population recovery in affected freshwater macroinvertebrates will occur within an acceptable time. The duration of the mesocosm studies limited to 147 days, after which autogenic recovery may have occurred. Additionally, the mesocosms were not connected to any non-contaminated stretches that may serve as sources for recolonization (Van den Brink and Baveco 2009). MASTEP was applied with the default parameterization for *Asellus aquaticus*, as presented in Van den Brink et al. (2007a). However, because risk has been identified also for glassworms, ephemeropterans and for other sensitive freshwater macroinvertebrates, conclusions from the modelling study should cover also those species. Due to the specific parameterization for *A. aquaticus*, results cannot be readily transferred to other species with different potential for reproduction and dispersal. Therefore, it is difficult to assess the level of protection for other species that can be drawn from the modelling conclusions.

3.8.3 Supporting Data

Life history parameters in the basic population model of MASTEP have been parameterized using a thorough research on the literature available for *A. aquaticus* covering the last 50 years (Van den Brink et al. 2007a). MASTEP is a simplistic population model that requires only few parameters (see model

[36] Verboom, J., H. Baveco and P. J. Van den Brink (2005): A Simulation Model for Spatial Population Dynamics of Asellus aquaticus after a Spray Drift Event of Deltamethrin in Aquatic Ecosystems. Alterra Wageningen University and Research Center; PO Box 47, 6700 AA Wageningen, The Netherlands. Sponsor: Bayer Crop Science AG, Development - Ecotoxicology, 40789 Monheim, Germany.

description in section 2.5). The model has been parameterized for Central European climatic conditions, with most values derived from peer-reviewed field studies and from expert judgement. Pesticide properties of deltamethrin were parameterized using the mesocosm study from Heimbach et al. (2005, see below).

3.8.4 The Environmental Scenario

Exposure and effects of deltamethrin were simulated in three environments that were considered to represent the standard FOCUS scenarios for a Central European pond, ditch and stream, respectively. Simulations started at 1st of January (day 1) with 1,000 individuals randomly placed across the water cells in each scenario, but populations increased to 7,000 – 9,000 individuals in summer. Because the default parameterization for *A. aquaticus* in Central Europe was used (see the general model description in section 2.5), a minor reproduction peak occurred in the simulations by the end of April (around day 120), and a larger peak in July (around day 190). In the model, all individuals did not produce more than once. 1 % of daily movement steps of *A. aquaticus* was considered to result in downstream drift events. The stream and ditch scenarios differed only by flow velocity (250 m / day vs. 10 m / day) that affected the drift of individuals and pesticide exposure. Drift distance for individuals in the stream scenarios was modelled to follow an exponential distribution with an average distance of 10 m, but was not relevant in the ditch scenarios.

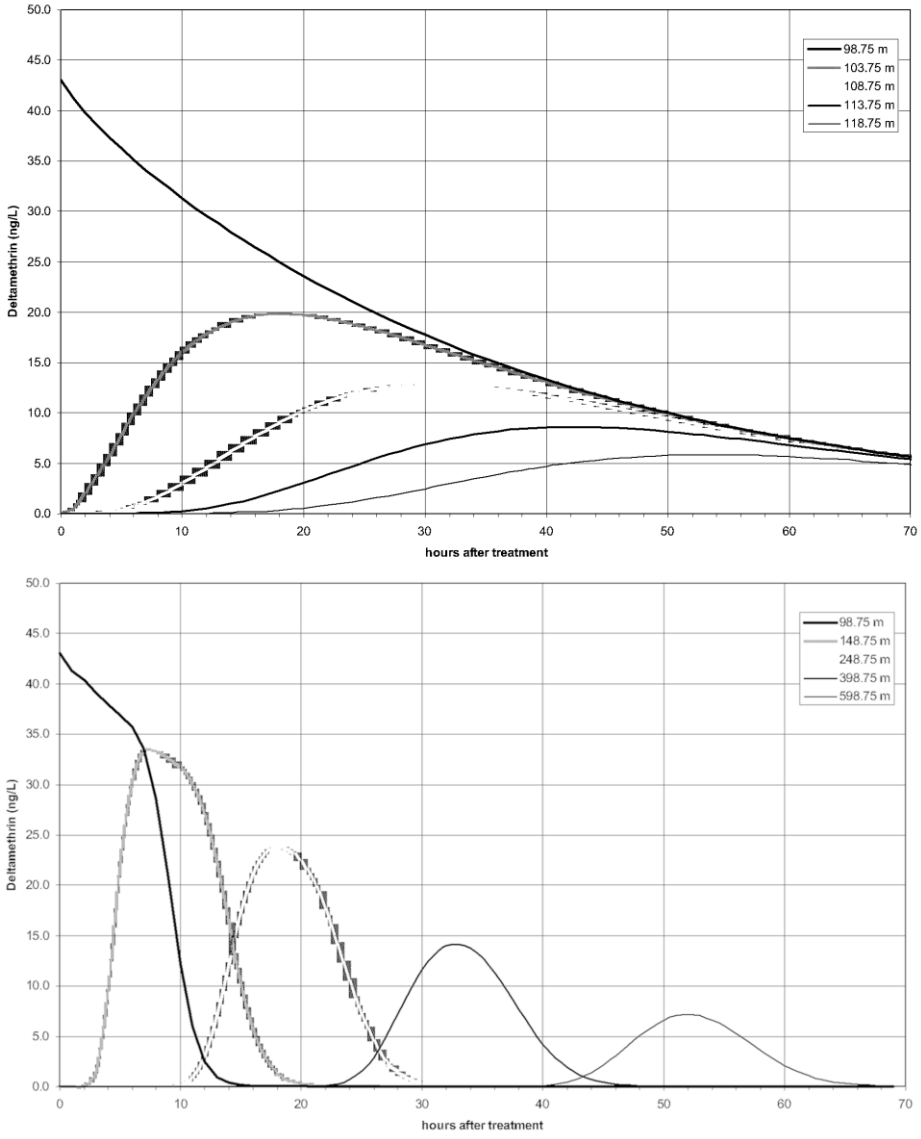
The landscape was modelled using a grid of quadratic cells with a size of 1 m^2. MASTEP was coupled with the fate model TOXSWA 1.2 for the prediction of pesticide loading that served as driving variable for MASTEP. For each environment, TOXSWA provided peak concentrations for the exposure via spray drift after application to different crops and with different buffer zones of 0, 5 and 20 m. Hence, initial peak concentrations of 16, 23, 30 and 43 ng/L were simulated. A single exposure peak on day 130 was simulated. However, the corresponding acute effects in MASTEP were scaled such that they matched the effects after a series of three consecutive pulses with an interval of 1 week, as expected following the proposed general application patterns for the reference product of deltamethrin (see parameterization below). The scenarios thus represented an application series in early May, during the first population peak. However, the proposed GAP for the reference product of deltamethrin allow also for late application in autumn after the second reproduction peak. Since the reproduction peaks are fundamental for population recovery and thus for the duration of pesticide effects at the population level, the early application in the modelled scenarios seems not representing a realistic worst case.

The pond was modelled using 30 x 30 water cells (of 1 m^2 each) and considered to neighbour an agricultural field treated with deltamethrin at one side. Therefore, at the day of exposure, cells that directly bordered the treated field were loaded with the full peak concentration of deltamethrin calculated from TOXSWA. Cells further away were loaded with amounts that decreased with distance from the edge of the treated field. The stream and ditch were modelled as a single row of 600 cells, of which the first 100 were directly exposed to deltamethrin. The end and beginning of the modelled stretch were connected, allowing individuals and the pesticide to move from the last to the first cell and vice versa (periodic boundary conditions). The built-in fate module calculated first-order kinetics dissipation and downstream drift of the pesticide. Therefore, the ditch and stream scenarios represented a 600 m stretch from a 1 m wide water body that borders a treated field for 100 m in every 600 m. Similarly, the pond scenario covered an in- and outflow represented by 39 cells connected at the borders of the landscape.

The landscape compositions outlined above may not reflect realistic worst-case assumptions in an area with intense agriculture, where non-treated habitats are rare. On the other hand, deltamethrin may not be applied to all fields at the same time, therefore the scenario may not be too optimistic when considering only the effects of this insecticide. However, from a scientific point of view it is a shortcoming that risk assessment of pesticides is conducted for individual active substances or products only, without considering the effects of additional substances that may be applied at the same

time. This general shortcoming in the governmental risk assessment procedures aggravates in a refined risk assessment as performed in this study, where recolonization due to spatiotemporal variation in exposure is considered without also considering effects of additional pesticides. In a landscape with intense agriculture deltamethrin may not be applied in all fields at the same time, but recolonization may be hindered due to the effects of additional pesticides applied in the remaining area at unknown dates. Life history parameters in MASTEP have been parameterized with field studies on populations that have likely not been heavily exposed to agricultural pesticides, therefore the effects of additional pesticides are not implicitly covered in the model. In MASTEP, effects of additional pesticides might be considered by modifying background mortality (and probably reproduction and dispersal) in the treatment scenarios. However, this would require convention on the magnitude and timing of such additional effects.

Figure 45: Application of MASTEP to Deltamethrin – Predicted Spatiotemporal Exposure



Simulated exposure profiles after pesticide application in different stretches of the ditch (top) and stream (bottom) scenario. With increasing distance after the end of the stretch neighbouring the treated field (first 100 m), the pesticide peak occurs with a delay due to drift (depending on flow velocity) and reduced height due to dissipation while drifting. Effects on local population size were modelled based on the peak concentration at the day when the peak occurred. While the population model was run in daily time steps, the fate module proceeded in smaller time steps. Graph reproduced from Verboom et al. (2005).

3.8.5 Parameter Estimation

MASTEP has been applied with the default parameterization for Central European populations of *A. aquaticus* as described in the general model description in section 2.6. Acute lethal effects of deltamethrin exposure on *A. aquaticus* were parameterized using results of the mesocosm study from Heimbach et al. (2005). In this study, mesocosms with a depth of 1 m were exposed to nominal concentrations of 0, 4.8, 10.5, 23, 51 and 111 ng deltamethrin/L in the water phase. The design consisted of three replicate mesocosms for the control, one replicate for the highest concentration, and two replicates for all other concentrations (Verboom et al. 2005). Following the GAP, the pesticide was applied three times with an interval of 7 days, resulting in peak exposure at days 0, 7, and 14. 300 individuals of *A. aquaticus* were introduced to each mesocosm prior to the first exposure. Populations of *A. aquaticus* were monitored using artificial substrate samplers and leaf cages.

Acute effects in MASTEP were based on a logistic regression of population decline (difference to population size in the controls) observed 21 days after first exposure vs. the nominal concentrations. The authors considered this approach valid because it facilitates direct coupling of nominal concentrations and decrease in population sizes without the need of modelling the pesticide's fate within the mesocosms in detail (Verboom et al. 2005). Measured concentrations reached on average 94 % of the nominal concentration after 4 h and then declined rapidly within 24 h, indicating that deltamethrin is a fast-acting, fast-dissipating insecticide and thus matched the general domain of applicability of MASTEP. However, vertical stratification of deltamethrin was observed after application, with only ca. 20 % of the nominal concentration expected at the bottom of the water body where *A. aquaticus* is living. Risk assessors considered this distinct stratification not representative for natural water bodies in which higher degrees of mixing due to wind and water flow were expected (UBA 2010). On the other hand, wind may increase drifting of macrophytes and thus recolonization of associated *A. aquaticus* (Verboom et al. 2005). In conclusion, limitations in the methodology of the mesocosm study used for parameterization may have resulted in an underestimation or overestimation of acute effects simulated in MASTEP.

In MASTEP, the full reduction in population size observed at day 21 after a series of three consecutive pesticide applications was executed at a single day (first day of the application series). The authors considered 21 d after the first treatment (7 d after the last treatment) a good timing for the assessment of effects on the population size to construct the dose-response curve in MASTEP. Data on later days were not available due to the introduction of new individuals in the mesocosms at day 21. According to the authors, assessing the population size in the samplers earlier (shortly after exposure) might have resulted in an overestimation of lethal effects due to transient immobilization that hinders individuals to move to the samplers (Verboom et al. 2005). However, immobilization due to pesticides is likely associated with effects on reproduction (and possibly survival) later on in life history, so that immobilized individuals in acute tests may be considered "ecologically dead". Since sublethal effects were not considered in MASTEP, it seems more realistic worst-case to consider immobilized individuals as dead when parameterizing acute effects. The mortality data used for parameterization of effects in MASTEP may cover also some indirectly induced effects such as an increased risk of predation and starvation due to depletion of energetic resources in the first days after exposure. However, populations in the mesocosm likely experienced lower levels of environmental stress than those in natural water bodies (e. g. no current or predation by fish in the mesocosms), therefore such indirectly induced mortality in the field may has been underestimated (Verboom et al. 2005).

Finally, the fitting of a single dose-response relationship across all age or size classes may have resulted in an underestimation of effects for young individuals which are particularly relevant for population recovery (Verboom et al. 2005). The authors of the modelling study investigated the mesocosm data to identify whether small and large individuals were markedly different in their sensitivity to deltamethrin, i. e. whether a single dose-response relationship for all age classes was acceptable. They hypothesized that if young individuals were more sensitive, the fraction of small individuals will increase

in the treated samples compared to samples from the control mesocosms. Therefore, the authors performed pairwise comparisons of the fraction of young individuals observed in the different control and treatment mesocosms at each measurement day. They concluded that the sensitivity of young and old individuals did not differ remarkably because overall, the fraction of small individuals was not lower than in the control in significantly more than 50 % of the pairwise comparisons. However, this interpretation does not consider that in the lower concentrations (4.8 and 10.5 ng/L), the proportion of small individuals was actually higher than in the control in most of the comparisons. This proportional increase in small individuals at low test concentrations reversed at high concentrations, in which a proportional decrease of young individuals was actually observed twice as often as a proportional increase (see Table 3 in Verboom et al. 2005). This pattern suggests that exposure to low concentrations may have induced premature release of young individuals as part of a stress response in mothers (emergency reproduction), as described e. g. for *Daphnia magna* (Kooijman 2000, p. 232). Additionally, low concentrations may have been lethal to senescent large individuals; senescent and young individuals are often most sensitive to toxicant stress. Higher concentrations appeared to be lethal to young individuals, explaining the observed dose-dependent reversal in the effects of deltamethrin on population structure. The mesocosm data therefore suggest that young individuals may be more sensitive to the higher test concentrations (23 – 51 ng/L).

3.8.6 Sensitivity and Uncertainty Analysis

A thorough sensitivity or uncertainty analysis has not been presented for the applied regulatory models. By the time when the study was submitted to be used in assessment, there was also no sensitivity analysis available for MASTEP in general.

In order to assess the importance of dispersal with drift in the stream scenarios, the scenarios were repeated without drift of individuals. In the 100 m treated section, predictions without drift differed from those with drift and showed no population recovery during the simulation period of one year. The overall population size in the total 600 m stretch, however, did not change significantly when drift was switched off. Means and 95 % confidence intervals were calculated from five model runs for each scenario. In the pond scenario, confidence intervals for the population size were small (< 10 % of the mean), whereas for the treated stretch in the ditch and stream scenario, confidence intervals were considerably larger (30 %).

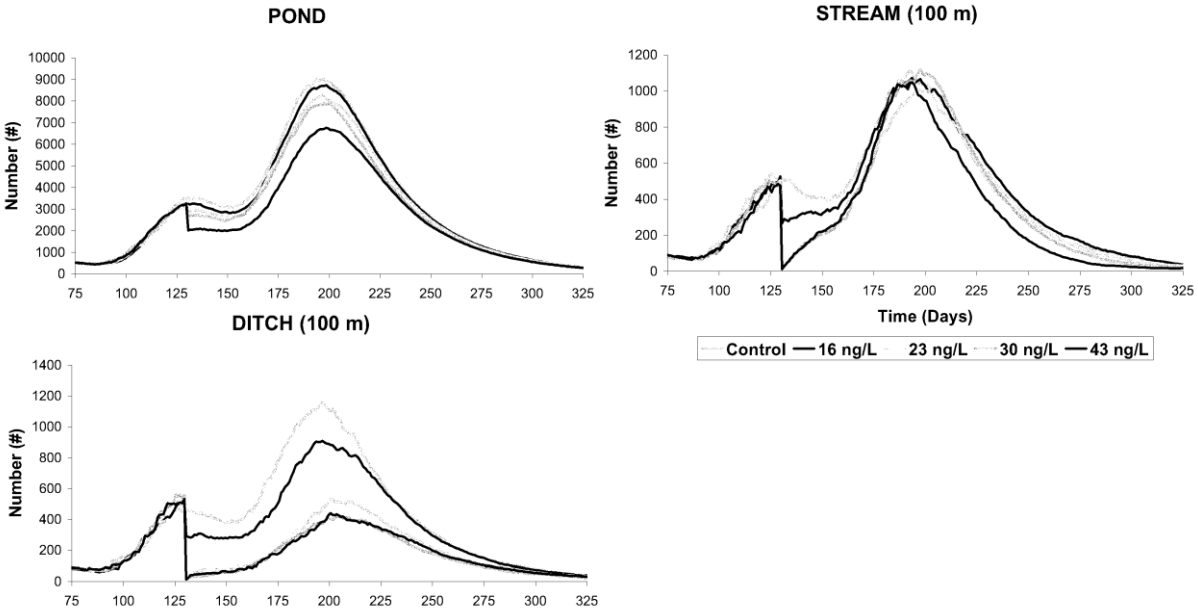
3.8.7 Comparison with Measurements

Predictions of MASTEP on population dynamics without pesticide exposure and on population recovery after pesticide exposure have not been subjected to validation prior to the application of the model in risk assessment by Verboom et al. (2005). Also in this study, the simulated population recovery could not be directly compared with the observed recovery in the mesocosm study, because new individuals have been introduced to the mesocosms 21 days after the first exposure, and because non-treated section were not present in the mesocosms. MASTEP is a simplistic model, omitting or simplifying various processes and factors; e. g., the environment is considered constant and energy budgets (leading to potential sublethal effects) or species interactions are not explicitly simulated (Verboom et al. 2005). The authors consider model results robust and hypothesize that the inclusion of additional mechanism will make the model more complex but not necessarily qualitatively better. However, sublethal effects of pesticides on reproduction and dispersal are likely to occur (particularly after exposure to concentrations that cause acute lethal effects) and may severely affect population recovery. Given the level of structural uncertainty and the uncertainty in parameterization (see above and the general model evaluation in section 2.6), the robustness and realism in model predictions is difficult to assess and should be tested prior to model application in risk assessment by a thorough comparison with independent data.

3.8.8 Model Use

In summary, effects of deltamethrin exposure in all simulations caused a concentration-dependent decrease in population size in contaminated parts of the landscape. The effect was most pronounced during the population peak in summer and diminished with the general population decline in autumn (Fig. 43). Strongest decrease was observed in the ditch scenario. The authors explained this pattern with fast recolonization due to drift in the stream scenario; also in ponds recolonization was facilitated by the two-dimensional structure of the water body. In contrast, recovery in the ditch was slow, illustrating the sensitivity of predicted population recovery to drift in MASTEP. When drift of individuals was switched off in the stream scenario, effects in the stream were comparable to those in the ditch.

Figure 46: Application of MASTEP to Deltamethrin – Simulations for All Scenarios



Predicted population sizes for the different freshwater bodies and initial peak concentrations of deltamethrin. The vertical axis shows the mean population size in the treated 100 m section (ditch and stream) or in the total pond across 5 model runs. Graphs reproduced from Verboom et al. (2005).

The authors reported that effects on the overall population size in the full simulated ditch and stream were lower than those in the 100 m treated stretch, but no figures were presented for the overall population size. According to the description, the overall effects were largest in the stream scenarios because drifting of deltamethrin resulted in high peak exposure of most parts of the modelled water body, whereas exposure in the pond and ditch remained spatially limited. Hence, the lowest peak concentration (16 ng/L) immediately reduced the population size in the treated 100 m stretch of the ditch and stream by ca. 30 %, but had almost no effect on the overall population size in the pond (Fig. 43). Populations in the first 100 m of the ditch took ca. 95 days (around day 225) to recover, when 95 % confidence intervals of their population size fully matched those of the control (Apendix 5, Figure 1 in Verboom et al. 2005). Confidence intervals in the stream overlapped after only ca. 20 days (around day 150), followed by a slight overshoot in population size in autumn. After exposure to higher concentrations, only few individuals in the treated stretch survived and confidence intervals did not overlap until the end of the simulation after 1 year both in the stream and ditch scenarios.

3.8.9 Overall Judgement

Overall, the application of MASTEP for the risk assessment of deltamethrin was generally well documented. The modelling results are specific for *A. aquaticus*, and conclusions for other, potentially more

vulnerable species should be drawn with great care. The authors of the study identified a need to improve parameterization related to density dependence and dispersal. Nevertheless, they considered the modelling results robust and representative for a realistic worst-case scenario due to a balance of assumptions that may result in an over- or underestimation of recovery (Verboom et al. 2005).

The results show that MASTEP is capable of producing long-term population decline after high acute effects and suggest that *A. aquaticus* can recover within the same year after exposure to 16 ng/L deltamethrin, but not after exposure to higher concentrations. However, the model predictions are associated with considerable uncertainty that extends beyond those depicted by the 95 % confidence intervals. Relevant sources of uncertainty include uncertainty in the parameterization of the life history and dispersal, but also of the acute effects of deltamethrin.

The study revealed considerable potential for an underestimation of pesticide risk: E. g., parameterization with effects after 28 d covers no transient immobilization or long-term delayed effects; stratification of deltamethrin in the mesocosms may have limited effects used for parameterization; no sublethal effects were considered; no indirectly induced effects via the community context were considered, such as a reduction in interspecific competitiveness or in the escape from fish predation; a single dose-response curve for all life stages was used. In comparison, potential for the overestimation of the risk of deltamethrin in the study (e g. no wind-induced recolonization) appears lower. Probably the most important potential for the underestimation of risk is the potential sensitivity of the simulation to sublethal effects on reproduction that have not been considered in this study. Even when the exact magnitude and duration of such effects is not known, it would have been interesting to run additional simulations with various degrees of reduction in reproduction after exposure. This way the potential influence of this important factor on modelling results may be assessed and better conclusions on the margin of safety can be drawn from the modelling study.

By the time of the model application, MASTEP had not been published to the scientific community yet (first publication in Van den Brink et al. 2007a), and no sensitivity analysis or validation of model predictions with independent data had been available. Therefore more scientific work on MASTEP may have been published prior to its proposed application for the risk assessment of deltamethrin in Verboom et al. (2005).

3.9 eVole – Application to Folpet

Evaluation by Jeremias Becker and Mathias Franz

3.9.1 Background

Folpet (*N*-(Trichlormethylthio)phthalimid) is a fungicide of the phthalimide family. During the chronic risk assessment of the combination product Melody® Combi, this active substance raised concerns for small herbivorous mammals when used in vineyards. Specifically, a study with constant high-dose exposure on rabbits (Rubin 1985) showed effects on the body weight of pregnant females, a reduced litter size and several developmental abnormalities. The study showed no acute lethal effects in the laboratory, but the observed sublethal effects may affect the performance under field conditions due to malformations and reduced energetic reserves, and suggest a potential for population decline. To demonstrate that there is no unacceptable risk from the proposed application, ADAMA Makhteshim Ltd. submitted a simulation study of Bastiansen and Meli (2016)[37] to the German EU zonal Rapporteur Member State (zRMS) in which the population model eVole 3.0 (RIFCON 2018) was applied. The study was formally rejected because it was provided too late in the registration process so that other Member States had no opportunity to comment. Additionally, the zRMS raised some content-related issues, claiming that the model was not sufficiently validated before application, that the applied regulatory model was not subjected to a sensitivity analysis, and that the simulations did not consider off-crop exposure that may result from spray drift.

3.9.2 Problem definition

According to the European framework for the risk assessment of plant protection products, mammals must be generally protected both at the individual and the population level (EFSA 2009b, EFSA PPR 2010). For Higher Tier assessments on birds and mammals, the actual protection goal of "no visible mortality and no long-term repercussions for abundance and diversity" has been defined (EFSA 2009b). However, this formulation is not precise on how to deal with effects that exert no acute mortality due to intoxication in the laboratory, but which may indirectly increase mortality under more challenging field conditions because they affect the performance and fitness of individuals. In case that indirectly induced mortality is principally considered acceptable at the individual level, it must be demonstrated that these effects will not result in long-term effects at the population level.

Rubin (1985) observed sublethal effects of the formulated plant protection product folpet which may translate into an indirectly induced mortality. Folpet exposure decreased body weight in rabbits, and e. g. Oksanen et al. (2007) showed on bank voles that survival may decrease with body weight. Additionally, the results from Rubin (1985) suggested that folpet may affect vole populations through a decrease in reproduction. The simulation study of Bastiansen and Meli (2016) was therefore submitted to provide a line of evidence that individual-level effects of folpet will not result in a long-term population decline of common voles. The authors chose the common vole as model species because it is considered representative of the generic focal species "small herbivorous mammals" in the European framework for the risk assessment of plant protection products (EFSA 2009b). The species is characterized by short generation times and high reproductive output and thus not an ecologically vulnerable species. However, other real species covered by this generic focal species share similarly low vulnerability, therefore the choice of the model species seems appropriate. Importantly, no specific protection goal has been specified for the modelling study. From the context it becomes clear that no or only limited impact on the population is acceptable. However, while the authors concluded that they

[37] Bastiansen, F. and M. Meli (2016): Population modelling for the common vole to assess the potential effects following the application of folpet in vines. RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany. Report No. R1520157. Sponsor: ADAMA Makhteshim Ltd., P.O.Box 60, Industrial Zone Beer-Sheva, 84100, Israel. Sponsor Ref. No. R-36423.

identified no significant impact on vole populations in the modelling study, it remains unclear what would qualify as a significant impact.

3.9.3 Supporting Data

The population model eVole is generally well supported by field and laboratory data from the scientific literature, although a lack of data has been identified for the spatial movement behaviour and for some life history traits that have been derived from laboratory studies under artificial conditions (see the general model evaluation in section 2.7). The model has been parameterized for Central European climatic conditions, therefore Bastiansen and Meli (2016) used the default parameterization for their model application.

Exposure to folpet was parameterized using dietary data from the EFSA guidance document for the risk assessment for birds and mammals (EFSA 2009b). Calculation of the daily dietary dose (DDD) requires information on the body weight of individual voles which is not simulated in eVole. The authors used a constant body weight of 26.06 g and 26.62 g for males and females, respectively, according to a study on adult voles of Baláž (2010) in Central European modern agricultural landscapes. The values range between those from EFSA (25 g) and from Niethammer and Krapp (1982, 27.6 g) which are derived from smaller sample sizes. However, the use of fixed values for body weight may have underestimated exposure of young individuals in the simulations because according to the equations used, the DDD will increase with decreasing body weight.

Effects of folpet in the model were derived from lower tier assessments which appear not well suited for the intended modelling purpose, particularly due to the low resolution in time. The EU Peer Review report on folpet (EFSA 2009a) concluded on a NOEC of 1500 ppm (141 mg a.s./kg body weight/day) for the long-term risk assessment for mammals, based on a two generation study in rats (Rubin 1986). However, a teratology study on rabbits (Rubin 1985) used for the setting of the Acceptable Operator Exposure Level (AOEL) observed a lower NOAEL of 10 mg a.s./kg body weight/day. Higher exposure to folpet reduced the body weight and the litter size of pregnant females and caused several developmental abnormalities in the offspring. A study from Akhurst (2005) suggests that the observed high sensitivity to folpet is due to the rabbit-specific gastrointestinal system, but because no data were available for voles, the zRMS requested the use of rabbit data as a conservative estimate.

Though no lethal effects on rabbits were observed, the results of Rubin (1985) suggest that folpet may increase the mortality of pregnant voles and their pups in the field. Specifically, Oksanen et al. (2007) observed a relation of reduced body weight and survival in bank voles. Bastiansen and Meli (2016) used this relation to translate reduced body weight to increased mortality in their modelling study. However, the quality of both supporting studies was not discussed in the modelling report. In particular, it is not clear whether the study was run sufficiently long so that effect sizes reached an upper asymptote with time. Additionally, while Rubin (1985) reported effects on the body weight of pregnant females, Oksanen et al. (2007) investigated the effects of body weight at birth. Therefore, the translation of decreased body weight in rabbits to increased mortality in voles is associated with high uncertainty, though effects might be rather overestimated due to the projection from early to late life stages.

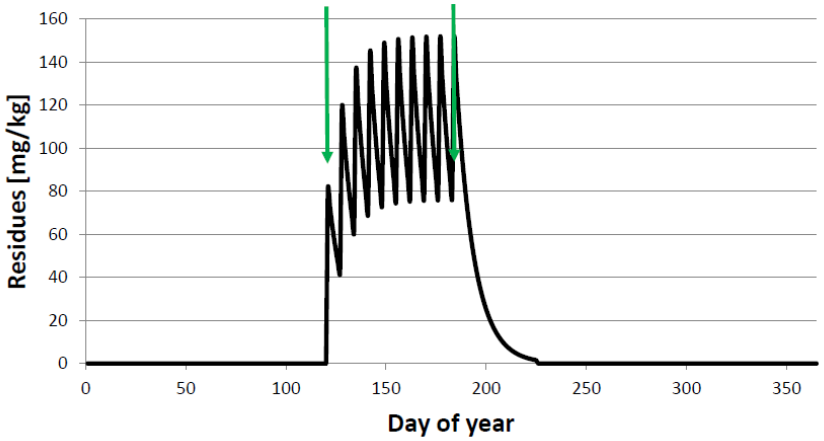
3.9.4 The Environmental Scenario

Simulations were run using a simplified landscape of 25 ha that consisted of two vineyards (75 % of the area) surrounded by grassland as off-crop habitat (25 %). This ratio of crop vs. off-crop habitat was considered conservative by the authors who stated that a typical agrarian landscape in Central Europe would probably contain more off-crop or other crop habitats. Additionally, the authors provided a summary on average holdings sizes in European wine regions, stating that two vineyards, each consisting of 9.375 ha surrounded by grassland, are representative. However, the data show that 24 % of the European holdings cover more than 30 ha, and a proportion of only 75 % vineyards in the landscape appears rather low in intensified wine-growing regions that represent a realistic worst-case.

In the simulations, the vineyards and off-crop grassland differed only in contamination, since voles were considered to use the grass between vines. It was conservatively assumed that the vole diet consists of 100 % grass, which provides higher residues than dicotyledonous plants (EFSA 2009b). Additionally, it was assumed that the full amount of applied folpet will reach the grass next to the vine (which is over-conservative because it would imply no efficacy of the treatment). On the other hand, only vineyard landscape cells were considered to be contaminated, though spray drift may also contaminate neighbouring habitats. Apart from folpet application, no additional agricultural practices ("landscape events") were explicitly simulated (but food was reduced twice per year to consider mowing events). On one hand, this may be a worst-case assumption because it makes on-crop and off-crop areas equally attractive to the simulated voles, while real voles might prefer undisturbed off-crop areas. On the other hand, no additional sources of mortality such as the destruction of nests from mowing are considered. The temporal dynamics in the availability of food and vegetation cover was set using mean vegetation data on Central European grassland from a series of three years by Jacob (2000).

Simulations were run for the earliest and latest possible folpet application scenario that followed the proposed general application pattern (GAP). The GAP implies an exposure to 10 × 1.5 kg folpet/ha with a minimum interval of 7 d between applications during BBCH stages 14 - 83 (in May – September). Therefore, an early application scenario starting at 1st May, and a late application scenario from 6th July to 7th September was simulated that covered effects during the early and late breeding season of common voles in Central Europe. For each scenario 50 "control" and 50 "treatment" replicate simulation runs were conducted, each pair of control and treatment scenario starting with the same initial conditions for better comparability. Simulations started at 1st of January and were run for 5 years pre-treatment, 10 years of treatment and 10 years after treatment (altogether 25 years). Each simulation covered on average > 20.000 individuals per year. Only dietary exposure was considered, though voles in vineyards may be additionally exposed through dermal contact during spraying. Additionally, only effects of folpet were simulated, leaving out potential effects from simultaneous exposure to the second active substance in the plant protection product (iprovalicarb).

Figure 47: Application of eVole to Folpet – Modelled Pesticide Applications and Residues in Food



Residues on plants are shown for the early application scenario using a DT50 of 6.22 days according to Knäbe (2013). Green arrows indicate the timing of the first and last application of the series. Graph reproduced from Bastiansen and Meli (2016).

3.9.5 Parameter Estimation

The general population model was run with the default parameterization for Central European vole populations as described in the general model description in section 2.7. Bastiansen and Meli (2016) provided a comprehensive documentation for the setting of all parameter values. The simulations

were initialized with a number of individuals drawn from a probability distribution that has been derived from test runs. In these test runs, eVole was initialized with a population density of 54 ± 2 individuals per ha, which corresponds to the typical density observed in the field according to the authors. Then, population densities on 1st of January from years 4 to 15 were recorded in the test runs, and a normal distribution was fitted to these data. Individuals were randomly distributed within the landscape and all started with a home range of 1 landscape cell. During a model pre-run of 14 d, the individuals then established their home ranges within d according to the rules of the population model, before the main simulation started. The initial age distribution was derived from an observed age distribution in the test runs after a simulation time of 1 year.

Exposure to folpet was calculated as the daily dietary dose (DDD) from intake of plant residues using the dietary equations provided in the EFSA guidance document for the risk assessment for birds and mammals (EFSA 2009b). In eVole, the intake of plant residues from the different landscape cells is proportional to the food values of the landscape cells that belong to a home range (see model description in section 2.7). Degradation of plant residues in each landscape cell followed single first-order kinetics. No random variability in the exposure of different vineyard landscape cells was implemented, which may have increased exposure in some individuals beyond critical effect thresholds. Accumulation of folpet in individuals was not considered, because the long-term study of Rubin (1985) implicitly covers potential accumulation if run for a sufficient amount of time so that exposure can reach equilibrium between uptake and elimination.

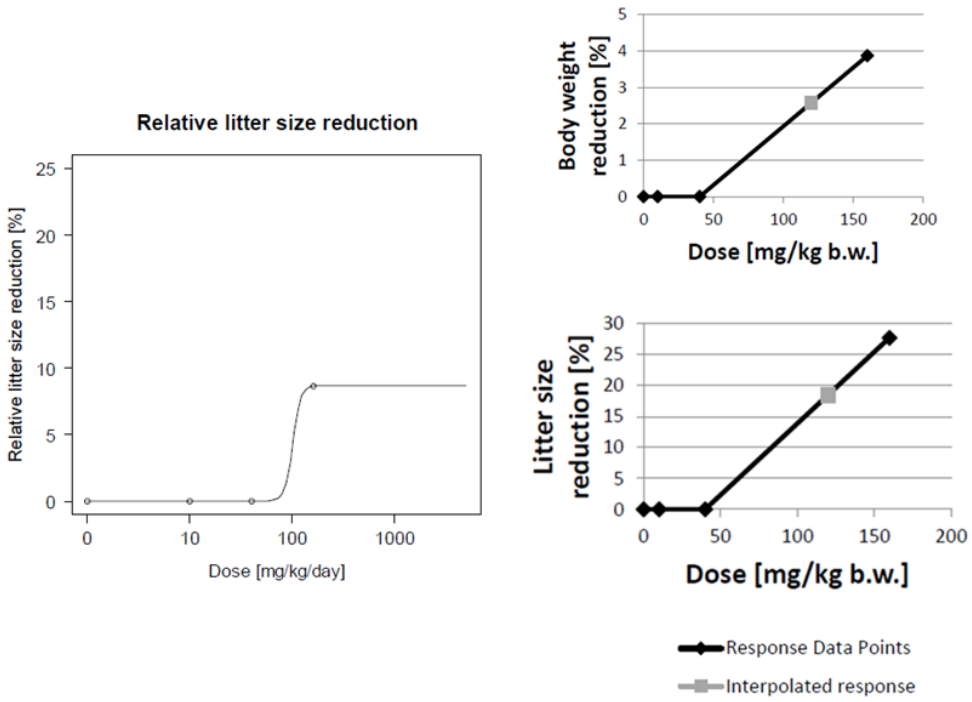
Effects were implemented on a daily basis using static dose-response curves instead of TKTD modules. Rubin (1985) exposed rabbits constantly to 0, 10, 40 and 160 mg a.s./kg body weight/day. Exposure to the high dose (160 mg a.s./kg body weight/day) significantly decreased the body weight of dams from day 4 of exposure until the end of the experiment. Exposure to the medium dose (40 mg a.s./kg body weight/day) slightly decreased the body weight of dams in the first 2 days of exposure, followed by a recovery from day 3 on. In order to conservatively use the maximum observed decrease in body weight for the modelling, the effect of the high dose at the last day of exposure (3.87 % decrease in body weight) and the effect of the medium dose at day 2 of exposure (0.25 % decrease) were used as endpoints. At 10 mg a.s./kg body weight/day, no effects on body weight were observed.

Exposure to the high dose decreased also the litter size through an increased post-implantation loss by 8.64 %. At lower doses no reduction in litter size was observed. Additionally, the following developmental effects on foetuses were detected by Rubin (1985): An increased probability for the development of a 13th lumbar rib was not considered relevant for modelling, because the presence of 13 lumbar ribs was also found in the majority of foetuses from the control group and considered part of natural variation with minimal effects on fitness. Also reduced and delayed ossification of parts of the tail bones was not considered relevant, because the tail of common voles is degenerated due to their subterranean life. Reduced or irregular ossification among sternebrae was not considered relevant because Collins et al. (1987) found that delayed ossification of sternebrae in embryos had nearly been reversed at day 6 post-partum and therefore may be transient and have no significant effects on further life history. In contrast, an increased risk of reduced ossification of the long bone epiphyses following exposure to the high dose was considered relevant for the modelling study because this effect potentially results in irreversible deformation of legs, with clear consequences for the fitness of individuals. It was pragmatically assumed that all affected individuals die before birth so that the effect was implemented in eVole as an additional reduction in litter size by 27.25 %. However, leg deformation may lead to death in later stages. In density-regulated populations such as those of common voles, the effect of premature mortality on population growth may increase the later it occurs in the life cycle: Affected individuals will compete for resources with non-affected individuals without contributing to reproduction. This limits the potential for population recovery due to competitive release after pesticide exposure. eVole is not able to capture this relationship because premature individuals

do not compete for resources. Setting the time of death to birth is therefore indeed the worst-case scenario that can be built in this model, because it reduces the population size at the earliest possible time point after exposure. However, modellers and risk assessors should be aware of the limitation of the model outlined above. At lower doses no developmental effects were observed.

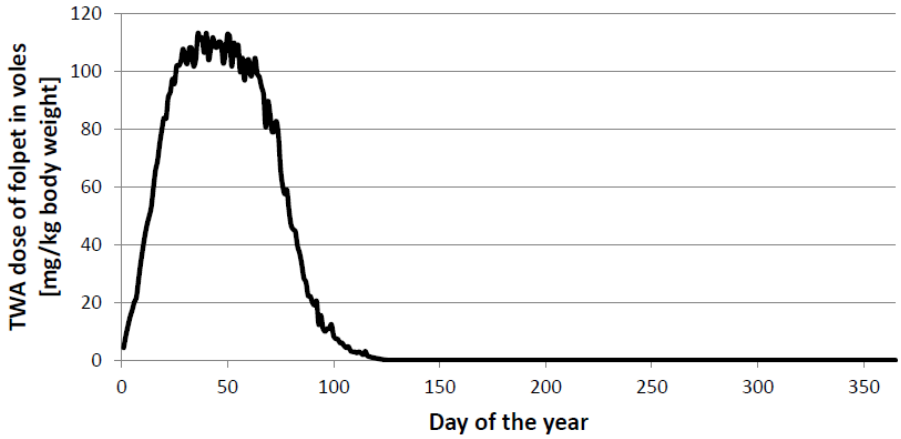
For the simulated effects of folpet, dose-response functions were established. A log-logistic dose-response curve was established for the direct effect on litter size (Fig. 48). Because an effect on litter size was observed only at the highest test dose in Rubin (1985), the upper asymptote for the maximum effect size could not be reliably estimated, and with the limited amount of data points the R function used for fitting can run into numerical problems without providing warnings. However, the exposure calculated in the eVole simulations did not exceed the highest test dose (Fig. 49), so that extrapolation was not an issue. For the decrease in maternal body weight and the decrease in litter size due to developmental malfunctions, no significant slope parameter could be estimated using log-logistic regression, and linear interpolation was used instead.

Figure 48: Application of eVole to Folpet – Parameterization of Individual-Level Effects



A log-logistic dose-response curve was used for the direct effect of folpet on litter-size. A linear interpolation was used for the effect of folpet on body weight in pregnant females (left) and for the indirectly induced effect on litter size due to developmental malfunctions of foetuses (right). Black data points show experimental observations. The grey data point shows the interpolated response for an exemplary dose of 120 mg/kg. Graphs reproduced from Bastiansen and Meli (2016).

Figure 49: Application of eVole to Folpet – Modelled TWA Exposure



Modelled temporal course of mean time-weighted average (TWA) dose of folpet across all exposed individuals. 10 applications were simulated between 1st January and 5th March. For demonstration purposes, the application pattern in this graph was set outside the reproduction period and does not match the GAP used in the model runs for risk assessment. Graph reproduced from Bastiansen and Meli (2016).

However, these dose-response functions relate observed effects to constant exposure over long time spans, whereas the population model proceeds in daily time steps. Therefore, in eVole the dose-response functions were not used to translate the exposure of the current simulation step (day) to an effect. Instead, each day a time-weighted average (TWA) dose of the last 21 days was calculated and applied to the dose-response function (Fig. 49). The approach assumes that at constant exposure, effects increase linearly over time until they reach the maximum. This is not a worst-case assumption, because effects may build up even in a short time of exposure (Liess and Schulz 1996) and may not increase much further under extended exposure due to repair and adaptation. E. g., the transient effect of the medium folpet dose on body weight in Rubin (1985) suggests such a temporal effect pattern. Using a TWA approach thus may underestimate effects of short-term peak exposure that can result from pesticide application and subsequent degradation of residues on plants. Additionally, for newborn pups, the TWA dose was calculated assuming no exposure before birth, although limited exposure in the womb may have already occurred.

The effect on litter size was executed at the last day of pregnancy (even if there was no ongoing exposure anymore). Each pup of the litter was then subjected to a common probability of death that is generated from the dose-response function, based on the TWA of the entire pregnancy period. In contrast, the decrease in body weight was translated to a decrease in daily survival using a relation in the bank vole observed by Oksanen et al. (2007). Because Oksanen et al. (2007) reported effects on survival 160 days after the measurement of body weight, the effect from this relationship was converted to reduced daily survival by taking its 160th root, assuming that decreased initial body weight affected survival in a similar way during each of the 160 days of the study. Again, this is not a worst-case assumption, because reduced body weight might have affected survival mainly during the first days, with little or no effects in later life stages. In that case, taking the 160th root may have underestimated the effect size of body weight on survival.

The implementation of effects from folpet exposure did not consider potential interactions with effects from additional stressors such as starvation or exposure to the second active substance of the plant protection product. Under laboratory conditions, individuals may have been able to partly compensate effects but may exert higher sensitivity under more challenging field conditions.

3.9.6 Sensitivity and Uncertainty Analysis

A sensitivity or uncertainty analysis has not been presented for the applied regulatory model. The authors referred to a previously conducted sensitivity analysis of the baseline population model (see the general model description in section 2.7). However, this analysis did not include effects of folpet or of chemicals in general. Therefore, the sensitivity of the modelling endpoint used for risk assessment (change in population size due to folpet exposure) to the substantial uncertainty associated with the implementation and parameterization of individual-level exposure and effects cannot be assessed.

3.9.7 Comparison with Measurements

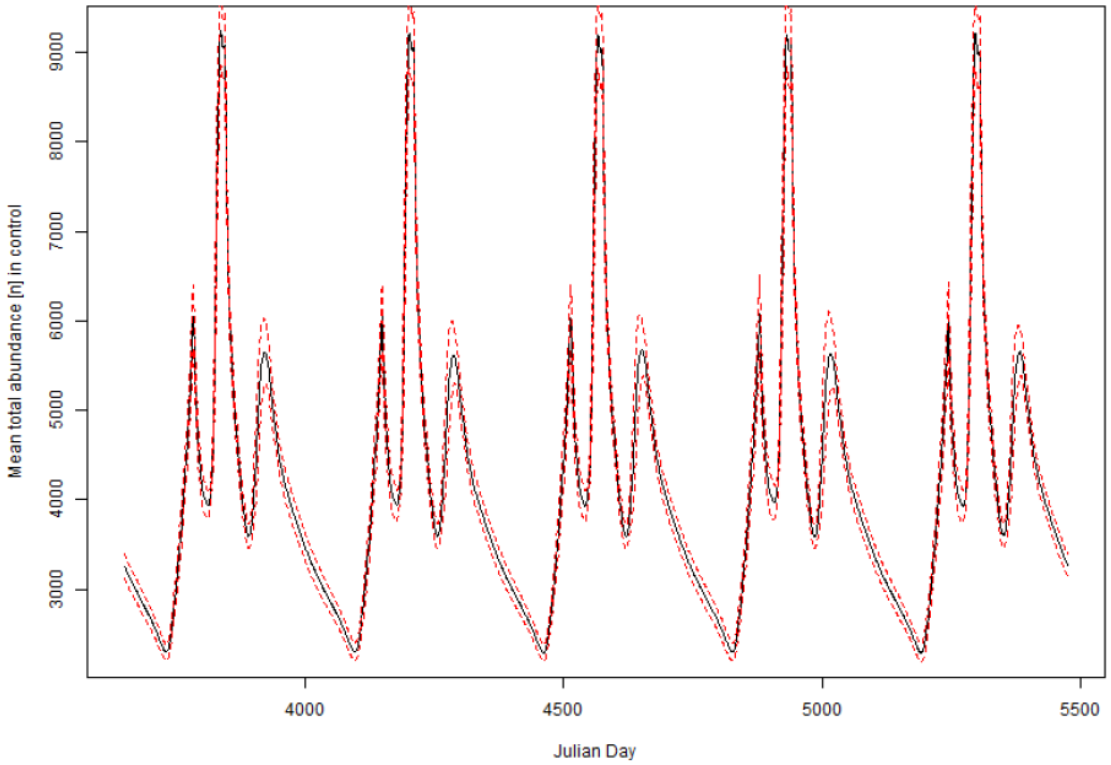
At the time when Bastiansen and Meli (2016) submitted the modelling study for risk assessment, reasonable matching of simulated population dynamics in eVole without pesticide exposure (incl. population size, reproduction, survival, spatial behaviour and age structure) with field observations had been demonstrated (see general model evaluation in section 2.7). In contrast, the ability of the population model to adequately predict population recovery or long-term population decline following pesticide exposure has never been tested. Additional stressors may limit the compensation of effects on population size or reproduction due to density regulation, even when dynamics in undisturbed populations may have been predicted well. Moreover, individual-level effects of folpet have been implemented with data from constant exposure under laboratory conditions; it has not been tested whether assumptions of the toxicity module hold also under time-varying exposure (see a discussion in Miller et al. 2000) or in the presence of additional stressors typical for field conditions that may increase the sensitivity of individuals. Therefore, it cannot be assessed whether the simulations may have underestimated effects at the individual level or overestimated compensation of effects at the population level.

3.9.8 Model Use

In the control simulations, the overall population size showed a seasonal fluctuation between ca. 2,500 and 9,000 individuals (100 – 360 individuals per ha, Fig. 50). Three distinct population peaks were observed during the breeding season that represented three largely synchronized litters of adult females, each followed by a sharp decline that resulted probably from high background mortality of subadults. Populations declined steadily in winter and decreased a minimum size before the beginning of the next breeding season.

Effects of folpet were investigated at the first and last year of the treatment period, and at the first year after the end of the treatment (years 6, 15 and 16 of the simulations). The cumulative daily mortality over year 15 (average across all simulation runs) in treated populations was compared to those of control populations. The difference could hardly be visualized, showing that mortality from folpet exposure was much lower than background mortality in the simulations. This is not surprising because the TWA daily dietary dose over 21 d never exceeded 115 mg/kg body weight/day (Fig. 49), which resulted in a maximum reduction in body size of ca. 2 % and was converted to a maximum additional mortality of <1 % according to Oksanen et al. (2007). In contrast, Fig. 47 suggests that the maximum actual daily dietary dose was approximately 1.5 times as high as the maximum TWA daily dietary dose. If effects were related to the actual daily dietary dose, this would have led to a decrease in body weight of ca. 4 % and translated to an additional mortality of ca. 5 %. As discussed above, from the available data it cannot be excluded that effects on body weight may be driven by short-term exposure and might have been underestimated in the simulations. The same applies to the reduction in litter size which reached to <9 % (direct reduction) and 27.25 % (reduction due to developmental effects on pups) in the simulations according to the authors.

Figure 50: Application of eVole to Folpet – Simulations for Control Scenario



Development of simulated control populations in eVole. The solid black line depicts the mean population size and dotted red lines depict lower and upper 95 % confidence limits from 50 replicate model runs. Graph reproduced from Bastiansen and Meli (2016).

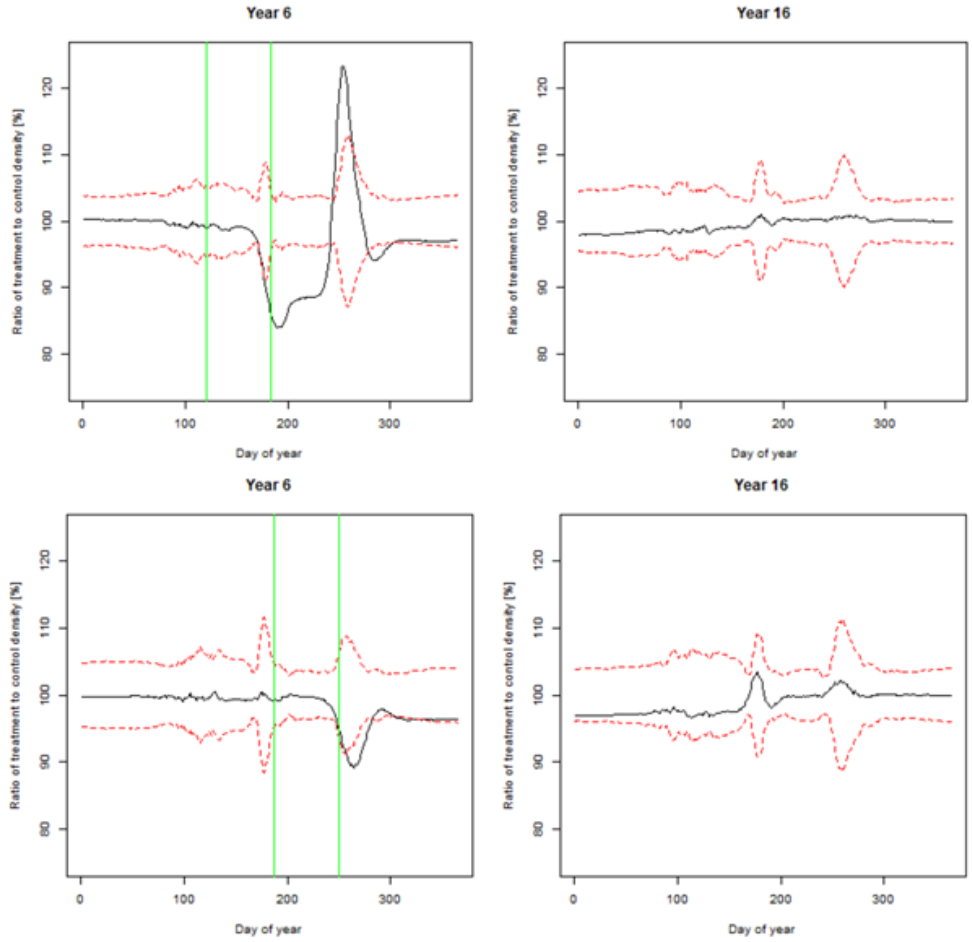
According to the authors, the maximum direct reduction in litter size was <9 % and the maximum additional reduction in litter size due to developmental effects was 27.25 % in the simulations. The influence of all simulated effects of folpet on population size were expressed as the relative deviation of the mean population size in treatment vs. control runs (Fig. 51). This type of documentation facilitates an efficient assessment of effect sizes. In the early application scenario, populations decreased to 83 % of the size of control populations towards the end of the application window. This was followed by an overshoot in population size to 120 % by the end of the breeding season, and a second but small decrease afterwards. The authors explained the observed overshoot with the mechanisms of density regulation implemented in eVole: Treated populations showed lower population density and thus a lower proportion of dispersing (homeless and thus non-breeding) individuals by the end of the application window. Consequently, reproduction in the following weeks was higher than in control populations. This was followed by a limitation of home ranges in the treated populations, so that reproduction decreased and the population size dropped again below those of control populations. In the late application scenario, population size decreased only to 87 % of control population size at the end of the application window. The relative population size almost recovered within a few weeks and showed no overshoot because no reproduction took place late in the year. In both scenarios, the average population size was 3 % lower than in the control runs by the end of the first year of exposure. The difference was within the natural variation in control runs and did not increase in the following years of application. When exposure stopped, the mean population size recovered back to those in the control runs during the following breeding season.

In control runs, the difference between the average number of births and deaths per year was close to zero. In the presence of folpet application, the number of deaths surpassed the number of births by 48

– 110 individuals per year. This number was small compared to the overall number of simulated individuals per year (> 20,000) and did not result in population decline across the 10 years of treatment.

Based on these results, the authors concluded that in the simulations folpet showed no significant impact on the population size of common voles. However, since no specific protection goals have been defined prior to the simulation study, it remains unclear what effect size might have been considered significant. In addition to effects on the population size, effects on the age structure might have been of interest, because structural endpoints may be more sensitive than cumulative endpoints.

Figure 51: Application of eVole to Folpet – Simulations for Application Scenarios



Modelled effects on vole population density in the early (top panels) and late (bottom panels) application scenario. The figures show the ratio of the mean population density across repeated treatment simulation runs to the mean density across repeated control simulation runs in the first year of application (left) and in the first year without exposure after continuous application for 10 years (right). 50 control and treatment simulations were run pairwise, with similar random parameter values for both simulations. Vertical green lines enclose the application window (first and last application of the yearly application series). The solid black line depicts the mean value, red dashed lines depict the 95 % confidence intervals for the control populations. Graph reproduced from Bastiansen and Meli (2016).

3.9.9 Overall Judgement

The regulatory model of Bastiansen and Meli (2016) for the risk assessment of folpet on common voles in vineyards was associated with high uncertainties that resulted mainly from insufficient toxicity data. Lower tier tests were used that have not been designed for the use in modelling studies. In particular, high risk of underestimating effects was identified due to the way how long-term individual-level effects observed under constant exposure were scaled to short-term effects under time-varying

exposure that could be used in the model. Additionally, potential effects of additional stressors in the field on the sensitivity of individuals to folpet, as well as on the population recovery via density-regulation were not considered. The arising uncertainties could be reduced by the use of ecotoxicological data obtained under varying exposure regimes that may allow the fitting of more advanced TKTD models for individual-level effects. Additionally, eVole should be subjected to a thorough validation of endpoints that are actually of interest for risk assessment, such as the potential of populations to compensate or recover from a given effect under field conditions. The case study of Bastiansen and Meli (2016) showed that populations in eVole almost fully compensate repeated reduction in litter size by up to 36 % (27.25 + 9 %) during the breeding season in a realistic landscape scenario for 10 years. A similar compensation ability of real populations is yet to be tested.

On the other hand, the simulation study included a number of assumptions that can be considered highly or even over-conservative. This includes the exposure of vole food to 100 % of the applied active substance, which implies no efficacy of the treatment on vine that is not eaten by voles (though potential exposure in off-field areas due to spray drift and volatilization / deposition was not considered). Additionally, voles were considered to feed exclusively on food sources that contain the highest residues (monocotyledonous plants). Despite the high overall uncertainty, the settings of the regulatory model therefore appear relatively balanced. However, this valuation should be supported with a sensitivity analysis to demonstrate that uncertainties leading to an underestimation of the risk will in fact not outweigh those uncertainties leading to a potential overestimation of the real risk.

Finally, an interpretation of the modelling results suffers from the lacking definition of a specific protection goal. The case study illustrates the need for the development of specific protection goals for common modelling applications in the framework of the European risk assessment of plant protection products.

4 Conclusions

Jeremias Becker with support of the whole consortium

After a short overview on ecological models that may be potentially of interest for the risk assessment of pesticides (plant protection products or active substances) in part 1 of this report, we described and evaluated 10 models in more detail in part 2. Most of these models have been already applied in dossiers for the registration of pesticides in the European Union, and we evaluated a number of such case studies in part 3. In the following, we discuss some general outcomes from the evaluations in parts 2 and 3 of this report.

4.1 Separation of General and Case-Specific Model Evaluation

We structured the description and evaluation of the general models according to the checklists for model presentation by the applicants and for model evaluation by the risk assessors as laid out in the EFSA Sci. Op. on GMP (EFSA PPR 2014b). Similarly, the specific model applications in part 3 of this report were evaluated based on the checklist for model evaluation from this document. For TKTD models that address effects of pesticides at the individual level, separate checklists have been developed in the EFSA Sci. Op. on TKTD Modelling (2018) but were not used in this report because they were available too late in the process.

However, the checklists for documentation and evaluation presented in the EFSA Sci. Op. on GMP (2014b) tend to mix aspects of both the development of a model in general and its application for a specific risk assessment. These aspects should be differentiated by using separate checklists for 1.) the documentation and evaluation of a model in general and 2.) of its application in a specific risk assessment. This way, applicants may refer to an existing documentation and previous evaluation of a regulatory model by risk assessors, and provide only a description and justification of the case-specific model application. This could improve consistency and reduce work load in model evaluation by risk assessors. The ultimate goal might be a comprehensive evaluation of commonly applied models by an EU agreed expert group (e. g. in EFSA) to establish a list of available regulatory models that are considered suitable for certain uses.

The checklists provided in the EFSA Sci. Op. on GMP (2014b) are structured according to the main steps in model development and use. These steps are often illustrated as part of a modelling cycle (see e. g. EFSA PPR 2014b, Grimm et al. 2014) and include the problem definition; supporting data; the conceptual, formal, and computer model; the environmental scenario; the parameter estimation; sensitivity and uncertainty analysis; comparison of model prediction with measurements (validation); and model use to address a given risk assessment question. While some of those steps can be clearly attributed to either the development of a model in general or a specific model application, others should be addressed in the description and evaluation of both.

When a model is evaluated in general, first it should be identified to which types of use the evaluation is limited, i. e. which risk assessment questions have been considered that might be addressed with the model (see section 4.2 below). When a specific model application is evaluated afterwards, it should be first identified whether the actual problem definition is covered by this domain of use of the model in general. Supporting data are required both for a model in general and additionally for a specific model application (typically information related to pesticide properties and use). .

Sections in the checklists on the conceptual, formal and computer model address aspects of the model in general. Evaluating a model in general may address the level of conservancy placed in the conceptual model based on the explicitly or implicitly covered model mechanisms and processes (see section 4.3).

Sections on the environmental scenario, the parameterization, sensitivity and uncertainty analyses, and on the validation of model predictions with independent data currently mix aspects of a model in general and of a specific model application which should be separated. Development of a model in general, at least in case of higher-level models for populations, communities and ecosystems, requires one or few default environmental scenarios for parameterization and testing. These default scenarios should be representative or even cover the intended domain of applicability, i. e. the range of possible scenarios to which the model is considered to be applied in risk assessment. E. g., a model could be developed and tested using scenarios with very high and low temperature to cover potential applications in the northern, central and southern EU zone. The environmental scenario used in a model application may then be adjusted to address specific conditions in a member state, e. g. in terms of field size and agricultural practices. Environmental scenarios in a specific model application should represent realistic worst cases at national or EU zone level, and should be covered by the domain of applicability of the general model (e. g., refer to the same EU zone). Additionally, the level of conservancy placed in the environmental scenario should be considered when evaluating a specific model application.

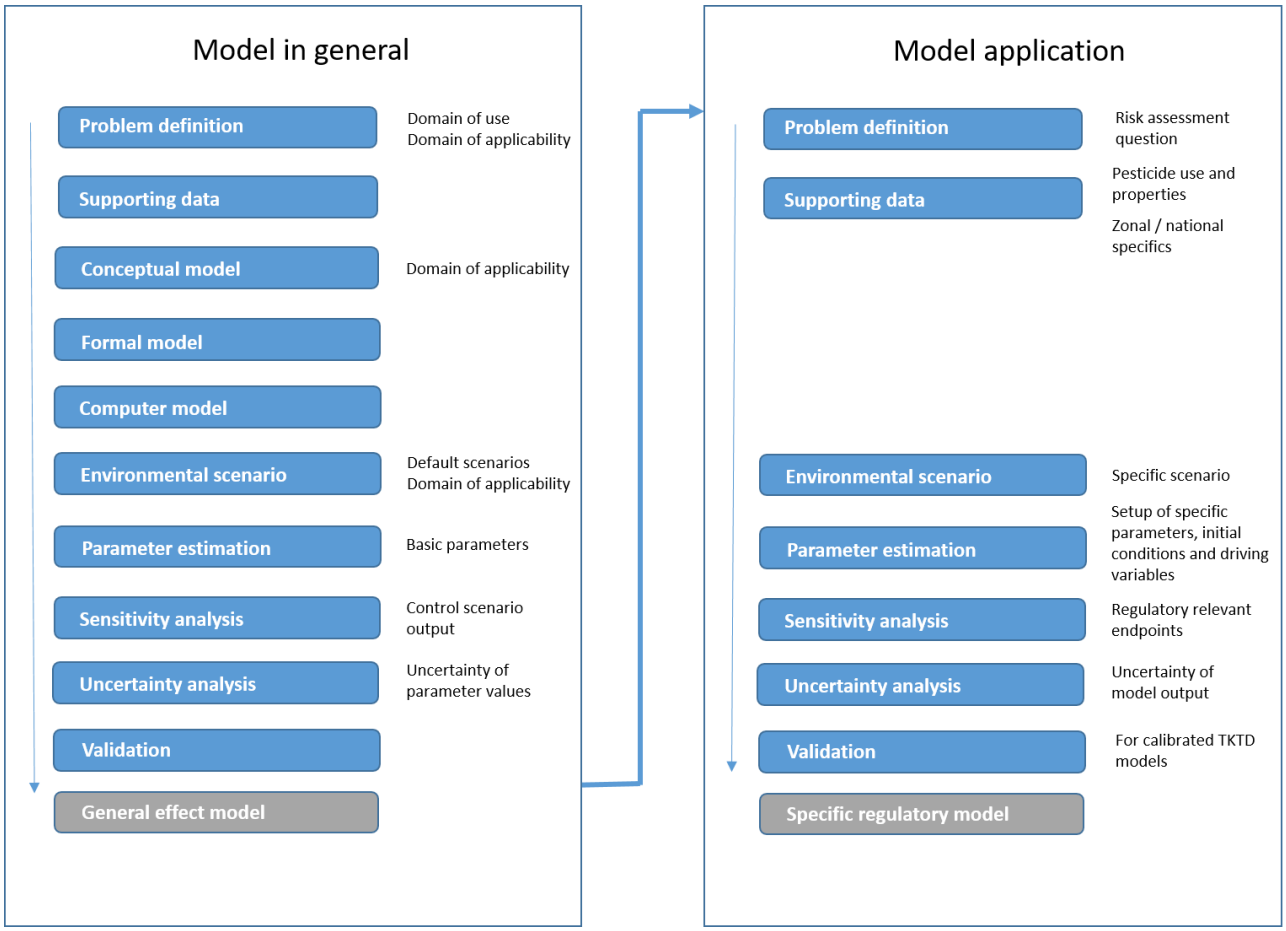
Parameterization of a model in general refers to values for principal biological (physiological and ecological) processes of a model and for the dependency of these processes on environmental conditions. Additionally, environmental constants that are typically not subject to change with the environmental scenario may be parameterized for a model in general. This basic parameterization constitutes to the domain of applicability of a model and may not be evaluated for each model application in a specific risk assessment. Changes may rather result in a new model version that would be need to re-evaluated. The parameterization of a model in general should be separated from the setup of this model for a given scenario in a case-specific model application. Model setup includes giving values to initial model conditions, and to parameters that are specific for a given environmental scenario. Additionally, model setup may include coupling of a model to a case-specific toxicity module, or at least the case-specific parameterization of an integrated toxicity module. With the toxicity module we refer to a non-mechanistical part in each effect model that converts pesticide exposure to input effects based on an empirical relationship. Similar to exposure models, the toxicity module may be coupled with (or integrated into) the main effect model, or may be run in advance and its output then be provided as input for a mechanistic effect model. The mechanistic effect model or model part then simulates the propagation of input effects provided by the toxicity module to output effects at a higher level of biological organization. E. g., in GUTS models, the toxicity module relates an internal pesticide concentration to an increase in sub-organismal damage; this propagates to mortality at the organism level. In DEB models, the toxicity module relates pesticide concentrations to changes in physiological process rates within organisms; this propagates to changes in survival, growth, maturation and reproduction at organism level. In these organism-level models, the relation of internal concentration with sub-organismal input effects is typically not accessible to direct observation; input effects are thus deduced from model calibration with empirical data on the organism-level output effects. In higher-level models, the toxicity module converts pesticide concentrations to effects on organisms (e. g. reduced survival, growth and reproduction). These input effects propagate then to effects on population size, community structure and even on abiotic conditions in case of ecosystem models. Toxicity modules of higher-level models are therefore typically directly parameterized with empirical observations from ecotoxicological tests. In all models, the selection (if not integrated) and parameterization of the toxicity module is subject to the case-specific evaluation of a model application.

Sensitivity analyses should be performed and evaluated both for a model in general and a specific model application, at least for those parameters that are not established by model fitting to a data set (see section 4.5). For a model in general, it is particularly interesting to assess the sensitivity of model output in a control scenario (such as temporal patterns in population development) to the basic model parameters (see above). For a model application, it is particularly relevant to assess the sensitivity of predicted pesticide effects in the model output to the specific parameters that describe the environmental scenario.

An uncertainty analysis may refer to uncertainty in model parameters or in model output (see section 4.9). Uncertainty in basic and in specific parameters should be addressed for a model in general and a specific model application, respectively. Uncertainty in model output due to error propagation should be assessed specifically in each model application.

Validation typically refers to a model in general, because the aim is to generate trust that a model predicts effects in a credible way across its domain of applicability, before it is applied in a specific risk assessment (EFSA PPR 2014b). As an exception, individual-level often models need to be validated in every model application (EFSA PPR 2018) because the input pesticide effects are obtained from case-specific calibration (see section 4.6).

Figure 52: Proposed Scheme for Model Evaluation



The figure presents a modification of the modelling cycle as presented in the EFSA Sci. Op. on good modelling practice (EFSA PPR 2014b). Blue labels illustrate the different steps in model development and application, together with the main aspects that may be evaluated by risk assessors. Graph provided by Jeremias Becker.

4.2 Model Applicability for Risk Assessment

The evaluated models address different levels of biological organization, ranging from the individual (GUTS, DEBtox) to the population (SpringSim, IDamP, MASTEP, IBM Chaoborus Population Model, ALMaSS, Wood Mouse Model of Liu et al. (2013), eVole), community ($SPEAR_{pesticides}$) and the ecosystem (AQUATOX).

4.2.1 Individual-level models

The GUTS and the DEB approach for individual-level effects cover a broad range of TKTD models for the prediction of acute mortality and of chronic (typically sublethal) effects, respectively. They may be used as an alternative to simple dose-response models for the fitting of Tier 1 toxicity data. In this case, the models may introduce additional complexity, but can handle repeated measurements and may thus provide fits with higher precision due to the use of more of the available data. However, in many cases repeated measurements are not available from Tier 1 data because they are not always required in the guidelines. To really optimize the potential of these models, guidelines for Tier 1 tests according to the data requirements (Commission Regulation (EU) No 283/2013 , No 284/2013 2013) may be extended accordingly.

The potentially greatest advantage of these TKTD models is, however, that they provide a method for extrapolation to different observation times and alternative exposure profiles that is underpinned with a mechanistic theory. Therefore, GUTS and DEB models may be most useful for the refinement of effects based on refined exposure scenarios as proposed in the Tier 2C approach on the aquatic risk assessment framework. Additionally, individual-level effect models may serve as toxicity modules in higher level models, i. e. building blocks that impose input effects of pesticides at organism-level (see section 4.1).

GUTS models typically require little information so that data from Tier 1 acute toxicity tests with repeated measurements are generally sufficient for model calibration. In contrast, DEB models require additional information on the physiology of the model species that varies with the complexity of the models. The simplest forms of DEB models can often be parameterized from the controls of a Tier 1 chronic toxicity test only, e. g. from a *Daphnia* reproduction test (Jager 2020). For the more complex "standard DEB model", an online-database (https://github.com/add-my-pet) has been established to collect such data, but the data quality is highly variable and there is a lack of case studies that successfully applied DEB models for risk assessment. Accordingly, in contrast to GUTS, the EFSA considered DEB models not yet ready for use in risk assessment (EFSA PPR 2018).

4.2.2 Population models

For population models, we identified principally two risk assessment questions that have been addressed in the case-studies published for model demonstration (see part 2 of this report) or proposed for regulatory risk assessment (part 3). First, population models have been used for the refinement of acute risk assessment, aiming at demonstrating that populations will recover from acute effects within an acceptable time period through reproduction and recolonization. Here we consider acute effects as "adverse effects of pesticide exposure occurring within a relatively short period after exposure", as defined in the EFSA guidance for aquatic risk assessment (EFSA PPR 2013). Second, population models have been used for the refinement of chronic risk assessment, aiming at demonstrating that chronic effects will not result in long-term repercussion on abundance. Chronic effects are considered as "adverse effects of pesticide exposure that develop slowly and / or have a long lasting course and that are caused by short- or long-term exposure" (EFSA PPR 2013).

4.2.2.1 Model application in acute risk assessment

Acute tests for Lower Tiers (Tier 1 and 2) report lethality or immobilization (which can be considered as "ecologically lethal" in a realistic environment) at the individual level within a short period of exposure to high concentrations (SANCO/10329/2002 rev. 2 2002, EFSA 2009b, EFSA PPR 2013). These effects are considered to result in an acute population decline. Additionally, an acute population decline may be directly observed in Higher Tier studies (see mesocosm example in section 3.7). According to the specific protection goals (SPG) of EFSA for terrestrial and aquatic invertebrates and plants, acute effects on the abundance or biomass of a population must be transient (EFSA PPR 2010,

EFSA PPR 2013, EFSA PPR 2014a, EFSA PPR 2015a): Negligible to small effects may be acceptable if they last for weeks to months, whereas larger effects may be acceptable only if they last no longer than few days. Population models for the refinement of acute risk assessment are thus used to predict the time period that a population requires for recovery from an observed or predicted decline. In the aquatic risk assessment, addressing the potential for population recovery is a refinement option in Higher Tiers termed the Ecological Recovery Option (ERO; EFSA PPR 2013). However, acute effects are generally not considered acceptable for vulnerable species characterized by ecological traits that suggest a low potential for population recovery, such as long generation times and low reproductive output (EFSA PPR 2010). Additionally, acute mortality is not considered acceptable for vertebrates that are protected at the individual level (EFSA 2009b, EFSA PPR 2010, EFSA PPR 2013). Therefore, population modelling for the refinement of acute risk assessment has been limited to fast growing aquatic and terrestrial invertebrates (see the examples of IDamP, IBM *Chaoborus* Population Model, MASTEP and SpringSim in part 2 and 3 of this report).

There is no general agreement on an exact definition of which effect size and duration is acceptable. As demonstrated in some modelling studies using MASTEP (Galic et al. 2012), recovery of an exposed local population may be faster than recovery of a whole metapopulation consisting of exposed and non-exposed subpopulations. Additionally, exposed populations may quickly increase back to the size of control populations, followed by an overshoot and potentially a second, smaller decrease afterwards, before population sizes match again. Finally, various criteria may be applied to identify when two population sizes actually match. These may be based on the magnitude or on the statistical significance of difference between the abundance of control and treatment populations (see discussion on model output presentation in section 4.8 below). Therefore, SPG should be established more precisely regarding the spatial scale (local vs. metapopulation), the temporal dynamics (overshoot and secondary decline) and the assessment methodology for population recovery. The limited specificity of protection goals can impede also the assessment of classical studies such as mesocosms, but becomes even more relevant for modelling studies which can provide much more detailed results for further analysis (i. e., there is almost no limitation on the number of replicates and observational time points in simulation studies, except due to computational effort).

Additionally, exposure to concentrations that cause acute mortality in some individuals causes also sublethal effects on growth and reproduction of the surviving individuals that typically persist for an extended period of time. Such chronic effects are usually not recorded in acute toxicity tests but are likely to hamper population recovery. Chronic lethal and sublethal effects from acute exposure may be also delayed and become visible weeks or months after exposure, when organisms enter a particularly sensitive life stage such as moulting or metamorphosis in insects (Liess and Schulz 1996, Beketov and Liess 2008b). Moreover, effects on the behaviour of organisms and their ability to cope with biotic and abiotic stress may be missed or strongly underestimated under artificial laboratory test conditions (see discussion on model design and parameterization in section 4.3 below). Finally, due to the exponential decline of chemicals in the environment, short-term exposure to high concentrations can be followed by extended exposure to low concentrations that may extend chronic effects due to chronic exposure. All these chronic effects limit the potential for population recovery and need to be considered when population models are used for the refinement of acute risk assessment. As standard acute tests do not provide this information, population modelling may be supported with chronic tests and additional non-standard tests. When chronic effects and effect interactions with the environment are not considered, the potential for population recovery is likely to be overestimated, and the aim of an increased realism in risk assessment will be missed.

In the reviewed case studies of population models for refined acute risk assessment, some sublethal chronic effects have been explicitly considered only in the application of IDamP to pirimicarb (section 3.5). Chronic mortality has been explicitly considered in the application of MASTEP to pyridalyl (sec-

tion 3.7). In contrast to the other case studies, effects in MASTEP have been parameterized with an observed population decline in mesocosms instead of laboratory data. Therefore, the applications of MASTEP to pyridalyl and deltamethrin (section 3.8) implicitly cover also sublethal effects and their interactions with other environmental stressors that may have contributed to the overall observed effect. However, the observed population decline has been implemented exclusively as an increased mortality during exposure in MASTEP; therefore, the population decline during exposure might have been overestimated, but the pace of population recovery after exposure (i. e., the most relevant modelling output for risk assessment) may have been overestimated. The application of the *Chaoborus* IBM Population Model to beta-Cyfluthrin (3.6) considered no chronic effects.

4.2.2.2 Model application in chronic risk assessment

Population models have been also used for the refinement of chronic risk assessment. These modelling studies aim at demonstrating that individual-level effects from the exposure to low doses (typically over an extended period of time) may not result in long-term repercussions on abundance / biomass of a population, as required in the SPG for animals and plants (EFSA PPR 2010). The studies often deal with vertebrates (small mammals in case of the application of eVole that has been reviewed as the only example in this report. Vertebrates are protected at the individual level, i. e. the SPG accept no lethal effects that could result in an acute population decline and would require subsequent population recovery (EFSA 2009b, EFSA PPR 2010, EFSA PPR 2013). The EFSA guidance for aquatic risk assessment describes a refinement option in Higher Tiers termed the Environmental Threshold Option (ETO) that aims at demonstrating that only negligible population effects will occur (ETO, EFSA PPR 2013). The ETO should be applied not only for vertebrates but also for invertebrates and primary producers if it is expected that highly sensitive species are affected that are ecologically vulnerable (see above).

As discussed for the definition of population recovery above, further agreement is also required on the exact definition of negligible population effects. E. g., long-term repercussions may be assessed by matching the average daily population size in a control and treatment scenario across several years. Alternatively, only the average population sizes during the seasonal reproduction peaks may be matched, which may provide a more sensitive measure of effects. Matching only the population sizes during the reproduction peak of the last study year may be even more sensitive because the effect accumulates over time. Additionally, as mentioned above, various criteria may be applied for the determination when two population sizes actually match (see the discussion in section 4.8 on the analysis of model output below). In conclusion, population models may directly address SPG in the framework of regulatory risk assessment, but illustrate the need for further refinement of these protection goals.

Studies in chronic risk assessment typically assess both lethal and sublethal effects that can be provided as input to population models. However, it should be considered that the input effects at individual level may be larger in a realistic environment that is more challenging as compared to standard tests conditions in the laboratory (see discussion in section 4.3).

4.2.2.3 Other potential uses

The potential application of population models in the environmental risk assessment of pesticides (ERA) is not limited to the use as refinement option in Higher Tiers. E. g., population models could be used to screen for potential worst-case scenarios to justify the setup of an experiment or field study. Additionally, population models may be used to assess the variability in population dynamics in non-exposed control populations across a high number of environmental scenarios. This information may be useful as reference for the development of SPG, as currently done by EFSA for honey bees using the BEEHAVE model (unpublished).

Apart from uses in prospective risk assessment, models for populations, communities and ecosystems may be also used for retrospective analyses of the impact of pesticides in the field. This has been the

prefered use of community and ecosystem models in ERA so far. Nevertheless, retrospective uses seem also promising for population models, because they provide the additional benefit of potential case studies for model validation (see section 4.6 below).

4.2.3 Community and ecosystem models

In the European framework for the risk assessment of pesticides, long-term repercussions for the diversity of non-target species are not considered acceptable (EFSA PPR 2010). Otherwise, no sSPG have been established for the community and ecosystem level. The assessment scheme follows the logic that if sufficient protection for a set of surrogate species has been established, this is considered to be protective also for the whole community and ecosystem (EFSA 2009b, EFSA PPR 2013, EFSA PPR 2017). However, it has been stated that Higher Tier assessments must integrate also the propagation of effects to the community-, ecosystem- and landscape-level (EFSA PPR 2013). Therefore, community and ecosystem models such as $SPEAR_{pesticides}$ and AQUATOX may be applied to back up Higher Tier studies on individual taxa. Additionally, these models provide the potential of assessing effects on a population in a more realistic environmental context that includes species interactions and environmental changes. Guidance for the risk assessment for freshwater organisms (EFSA PPR 2013) requires that conditions in Higher Tier test systems "are sufficiently representative of natural ecosystems in terms of species composition, species interactions (competition, predation) and natural stressors". This is generally not the case for environmental scenarios in population models for single species. Therefore, AQUATOX or other, potentially simpler community and ecosystem models may be used instead of simple population models to address SPG for the population level.

The SPEAR approach has been developed to quantify effects of pesticides on freshwater macroinvertebrates that can be observed at the community level in the field. SPEAR values relate to the proportion of taxa in a community that have been classified as ecologically vulnerable to pesticides. Because this information is of little help alone, SPEAR values have been empirically related to the estimated individual-level effects that have been experienced in the studied communities. Individual-level effects were assessed as toxic units, i. e. the pesticide concentration in the field vs. the LC50 of a representative species in acute toxicity tests. The established SPEAR vs. toxic unit relationship is currently used in a German national monitoring programme to study effects in the field, but could be also used to predict community effects from Tier 1 toxicity data.

SPEAR values may address the general or actual protection goal of no "long-term repercussions for the abundance and diversity of non-target species" (Regulation (EC) No 1107/2009 , EFSA PPR 2013). SPEAR values as the provided endpoint can address the the ecological threshold option (ETO), accepting negligible population effects only. However, they do not fit in the SPG established for freshwater macroinvertebrates that address the magnitude and duration of effects on the size or biomass of populations; the ecological recovery option (ERO) (EFSA PPR 2013). The relationship between toxic units and SPEAR values was established across a large number of pesticides with various chemical properties and modes of action. This relationship may not be customized for a specific application pattern and exposure profile of a compound to be assessed Therefore, in the current ERA framework, $SPEAR_{pesticides}$ may be applied for the retrospective ground-truthing (validation) of safety factors, as is currently done in the monitoring programme mentioned above. In prospective risk assessment, $SPEAR_{pesticides}$ may be used as an initial reference step: If a Higher Tier study identified considerably lower risk than those predicted from SPEAR modelling, this may trigger an extensive evaluation, because either indeed specific properties of the pesticide and its application may justify an unusually low risk, or relevant factors and processes may have been missed in the Higher Tier assessment.

4.3 Conceptual Model Design

Static, descriptive models such as dose-response or SSD models have been used in the regulatory risk assessment of pesticides for decades. In contrast, all models that have been evaluated in part 2 of this report, except of the $SPEAR_{pesticides}$ approach, are dynamic simulation models in which the output emerges mechanistically from the iterative use of a set of coupled equations or algorithms.

TKTD models for the effects of pesticides at the biological organization level of individuals are based on a set of coupled differential equations. The GUTS framework synthesizes a variety of relatively simple models for the prediction of acute mortality that require only little information for model fitting. In contrast, DEB models for chronic and sublethal effects require more information to parameterize, including information on a species' basic life history. The simplest DEB-based TKTD models can be parameterized using data from partial life-cycle studies (time-resolved observations on body size and reproduction), while the more complex "standard model" requires information from an online database. The evaluated TKTD models generally rely on a single compartment to represent the internal concentration or "damage" that drives the toxic effect. This is out of necessity, since in most cases measurements of internal concentrations over time are lacking in toxicity tests.

However, care should be taken because effects may be underestimated after transition to different life stages. E. g., individuals may start to mobilize contaminated reserves (separate compartment) that have been accumulated during the larval stages; this may explain strongly delayed effects (Liess and Schulz 1996) but is usually not considered in TKTD modelling. Additionally, TKTD simulations typically do not consider transfer of a certain level of contamination and effects (such as reduced energetic reserves) from the parents to their offspring. Care is thus needed when the model is extrapolated to cover untested life stages and generations.

All evaluated population models are individual-based models (IBMs), indicating a clear trend in ecotoxicological modelling for higher organization levels towards agent- or individual-based approaches. The main reasons for this trend may be an increasing computational power that used to limit the application of IBMs in the previous decades, and a potentially easier parameterization with data available from ecotoxicological studies.

An advantage of IBMs is that they predict population dynamics in discrete numbers, i. e. populations can actually drop to zero and go extinct without autochthonous recovery. In contrast, models based on ordinary differential equations (ODE) handle populations using floating values. Without some workarounds, populations in these models never reach fully zero and thus may always recover autochthonously when sufficient time is provided. Additionally, IBMs may provide more possibilities of considering demographic effects. E. g., effects on particularly sensitive life stages may increase synchronicity in the development of survivors that may increase competition and consequently delay population recovery through reproduction (Liess and Foit 2010b), However, this potential of IBMs has been addressed only to a small extent in the evaluated models and case studies, because many models do not distinguish life stages or consider only limited interactions between different life stages. Finally, IBMs facilitate the spatially explicit modelling of individuals that can move in a multidimensional environment, which is difficult to achieve with models based on differential equations. The evaluated population models for terrestrial organisms were always spatially explicit at an appropriate scale that matched the individual's presumed range of activity. In contrast, most models for freshwater organisms were not spatially explicit, possibly because the water column is considered more homogeneous as compared to the habitat of soil organisms or small mammals. As an exception, MASTEP provides a spatially explicit simulation of freshwater organisms, because the model focuses on source-sink (meta)population dynamics in the recolonization of contaminated water bodies. The non-spatial *Chaoborus* IBM Population Model incorporates recolonization via the use of simple migration rates for the non-explicitly simulated flying adults.

The level of detail that has been spent on recolonization processes in the population models (which potentially reduce population-level effects of pesticides) is generally contrasted by the exclusion of various processes that potentially increase effects. Populations in the field may be exposed to various sources of abiotic stress that may lower their potential to compensate or recover from the effects of pesticides. Such stressors may include e. g. desiccation, low levels of dissolved oxygen, heat stress or cold (Hardstone et al. 2009, Janssens and Stoks 2013, Russo et al. 2018), which have generally not been explicitly considered (however, see an example of dormancy in the IBM *Chaoborus* population model). In some cases, biological parameters have been parameterized with data on populations under (semi-)natural conditions, so that some level of additional stress has been considered implicitly. In other cases, the biological parameterization was done with data from artificial conditions that may not hold for real populations in the field.

Additionally, the design of the evaluated population models principally allows the inclusion of toxicity modules for both lethal and sublethal acute and chronic input effects at the organism level. Acute lethal effects are likely associated with chronic effects (e. g. on survival, growth and reproduction) that can affect the recovery of a population. However, chronic effects have been rarely implemented when population models were used to assess the potential for population recovery as a refinement in acute risk assessment (see discussion in section 4.2 above). Population models could be generally coupled to a variety of different types of toxicity modules. However, in many cases simple dose-response models were applied. The concepts of these modules are principally logical but require various assumptions on how effects observed at a single day may be extrapolated to different observation times and exposure profiles that may not always hold (see discussion on model validation below). TKTD models as toxicity modules could solve many of these issues and were used in some cases to model acute mortality (with different variants of the GUTS approach). However, we found only few and experimental applications of TKTD models for sublethal effects in higher-level models (DEB-IBM, see chapter 1), possibly because of the added complexity relative to a dose-response model.

Moreover, additional stressors such as competition and unfavourable temperature can increase the sensitivity of individuals to the effects of pesticides (Knillmann et al. 2012a, Russo et al. 2018). Similarly, pesticides can increase the sensitivity of individuals to additional stressors such as predation (Janssens and Stoks 2012), which can affect the potential of populations to compensate or recover from pesticide effects. Therefore, effects of pesticides on individuals may be difficult to detect in laboratory tests but become relevant in the field when they interact with additional stress from unfavourable abiotic conditions and antagonistic species (Liess et al. 2013, Becker and Liess 2015, Liess et al. 2016a). The toxicity modules used in the evaluated model applications do not explicitly capture these interactions of pesticide effects with additional stressors at the individual level. In MASTEP, pesticide-induced mortality was parameterized with data from mesocosm studies. The applications of MASTEP therefore implicitly covered a potential increase in mortality due to additional stressors in the field, as long as conditions in the mesocosms are considered as representative for conditions in the field (but see the limitations regarding chronic sublethal effects discussed in section 4.2 above). Additionally, the vulnerability of a population to pesticide effects can be affected by biotic stress from additional species such as predators (Berticat et al. 2004), competing species (Foit et al. 2012, Knillmann et al. 2012b, Becker and Liess 2015) and pathogens (Duron et al. 2006). A solution could be the use of extended population models: Some individual-based population models have been extended to two species to demonstrate the importance of species interactions on the modelling outcome (Gabsi et al. 2014d, Kattwinkel and Liess 2014). Yet, to our knowledge, such models have generally not been proposed for regulatory risk assessment. One exception is the application of IDamP to pirimicarb (section 3.5), where the population model for *D. magna* was coupled to a very simple ecosystem model with a single algal population as food source (StoLaM). This extended population model was called DaLaM.

Finally, regular population models cannot capture indirect effects on a model species that arise from effects on other species via the food web. This shortcoming can be addressed using community or ecosystem models. Another advantage of ecosystem models is the possibility of studying effects that may result from biomagnification. Most of these models are based on differential equations, probably because IBMs can become overly complex when a whole ecosystem is addressed. Mass-balancing models such as AQUATOX or CASM (Bartell et al. 1999) consider the flow of nutrients and toxicants through various compartments of an ecosystem which would not be feasible when simulated in discrete portions such as molecules. Instead, nutrients and toxicants are handled as concentrations in a compartment, and consequently biological compartments (populations) are described as the overall amount of biomass instead of individuals.

However, the appropriate parameterization to obtain a reliable prediction of such complex processes is very difficult. In AQUATOX, as a particularly complex and prominent example of ecosystem models that has been reviewed in detail, several default scenarios of different aquatic ecosystems are available. These scenarios have been fully parameterized with a large amount of ecological data, so that theoretically a user only needs to add toxicity information for relevant organism groups. However, there exists no agreement on a whole standardized community or ecosystem that could be modelled as representative for the field.

The toxicity module (termed ecotoxicology module in AQUATOX) is based on dose-response curves and was designed so that it can be used with only minimal toxicological information from Tier 1 tests. With a more refined approach for the direct input effects, it would probably not be possible to parameterize a whole ecosystem model. AQUATOX uses a complex methodology to circumvent the limitation of the dose-response approach to a fixed exposure time and to scale direct input effects to dynamic exposure profiles (see section 2.11). However, this requires a set of assumptions that may not always hold and represents an important source of uncertainty in the model. Therefore, the model has so far never been used for prospective risk assessment, but only for retrospective studies. Again, no interactions of the direct effects from pesticides and additional stressors at the organism level can be simulated.

As an exception, the application of the $SPEAR_{pesticides}$ approach for the prediction of community-level effects provides no dynamic simulation but a static linear regression of individual-level effects (toxic unit) vs. community-level effects (SPEAR value). When SPEAR is used as it was originally intended to, i. e. to quantify pesticide effects observed in the field, the SPEAR value emerges mechanistically from the ecological knowledge on traits that render taxa vulnerable or non-vulnerable to toxicants provided as model input. However, SPEAR values can be interpreted only on the base of the numerous field studies performed in various geographical regions. Therefore, the $SPEAR_{pesticides}$ approach provides an empirical approach for the linking of lower Tier individual-level effect towards the community-level effects. Such an empirical approach has the advantage that model predictions for effects in the field are generally realistic, whereas predictions of mechanistic models might be fully misleading to assess effects in the field as if relevant ecological mechanisms have not been included. However, accordingly the SPEAR approach is only customized to the acute effects of a specific pesticide (toxic unit) and not to its timing of application and exposure profile. Therefore, $SPEAR_{pesticides}$ modeling may be best applied as an initial step of risk assessment, with a low risk of substantial under- or overestimation of the real risk (see discussion on model applicability above).

4.4 Model Parameterization

Parameters can be broadly separated in basic parameters that are considered to be constant across different model applications, and specific parameters whose values vary with the environmental scenario and with the pesticide properties and use patterns in each model application (section 4.1).

In the reviewed individual-level models, many (and in some cases of GUTS even all) parameters values were obtained from calibrating the model to a specific data set in each application. As a consequence, individual-level models need to be fully or largely re-parameterized in each model application. Calibration in this sense means the fitting of all model parameters to produce the best fit for the data set at hand. This is sensible when a model includes only few parameters and when these parameters are not accessible to experimental observation, whereas the model output can be matched to available observations. E.g., mortality (model output) can be much better studied than the rate of an abstract damage built-up within the organism (model parameter) in case of GUTS models. Therefore, calibration (also called optimization) procedures receive ample attention in the literature on individual-level models.

In contrast, in case of higher-level models, observational data are available rather for the direct assessment of parameter values than for the matching of model predictions. E. g., it is typically easier to assess growth rates of individuals in laboratory or enclosure studies than population growth over several seasons. Additionally, many higher-level models include a high number of parameters, so that full model calibration would lead to considerable overfitting: with a high number of free parameters a model can be easily matched to closely reproduce a given output pattern, but the predictive power under different conditions will be low. Therefore, parameterization of higher-level models was mostly done with experimental data, and calibration was only used in few cases where empirical values were not available or could not be measured.

4.5 Sensitivity Analysis

The EFSA Sci. Op. on GMP (2014b) requires that models are extensively analysed before they may be used for the regulatory risk assessment of pesticides. However, though extensive general background information was provided in that document, the exact procedures required for sensitivity and uncertainty analyses and for the validation of models with observational data remain unclear.

A sensitivity analysis describes to what extent model output is affected by specific changes in model input (Loucks and van Beek 2017). In a local sensitivity analysis, one or few parameter values (or alternative sub-models for a specific model mechanism) at a time are varied in small and defined steps to assess the effect on one or several output variables. Varying an increasing number of parameters at the same time captures also interactions among parameters or sub-models (covariation) and leads to a global sensitivity analysis (Grimm et al. 2014). However, for complex models, a global analysis is often limited due to run time and difficulties in the statistical analysis of results. A sensitivity analyses is particularly useful to identify where uncertainty and variability is of highest concern in a model. For highly influential processes and parameters, even low levels of uncertainty or variability in the model input may lead to high levels of uncertainty or variability in the model output.

However, in TKTD models for individual-level effects, a sensitivity analysis may be of limited use because typically these models have been fully or largely calibrated to a specific data set (Jager and Ashauer 2018a). Therefore, it may be more informative to assess the confidence interval for each parameter estimate (and for correlations between parameters), as generated in the calibration procedure. Additionally, for models that have been fully calibrated to a specific data set, a sensitivity analysis would only make sense for this specific model application. Such models have no built-in basic parameter values and no default scenario (see section 4.1) that could be used for an analysis of the model in general.

In contrast, a sensitivity analysis may be highly informative to assess the performance of population models and of the ecosystem model AQUATOX. Because these models typically include a built-in basic parameterization for the biological part (see section 4.1), most sensitivity analyses may be performed for a model in general and not be repeated for each specific model application. Accordingly, for most of the models evaluated in detail, sensitivity analyses have been provided as part of the general model

descriptions in the open literature (see part 2 of this report). The model descriptions generally provided local sensitivity analyses for the default scenarios used for model demonstration. This may be acceptable, since a global sensitivity analysis covering the full range of possible parameter values and their combinations may be not feasible and not very informative due to its high complexity. However, sensitivity analyses of case-specific aspects, such as the selection and parameterization of a specific toxicity module for the input effects of a pesticide at organism-level, were often missing in the documentations of specific model applications for risk assessment.

In most cases, sensitivity analyses have been performed for the response of a number of output variables that may not be directly relevant for risk assessment, such as the average litter size or development time in control runs over one year. Assessing the sensitivity of such model predictions may be useful to assess the structural integrity of a model. If sub-models for different mechanisms such as development or reproduction are not overly sensitive, credibility increases that the model as a whole is robust against unforeseen conditions that might break the model. However, predicted pesticide effects as the typically most important model output from a regulatory pot of view was generally not subjected to a sensitivity analysis. Sensitivity of regulatory relevant model output to variation in model input (particularly in the regulatory relevant input effects and concentrations) should be assessed in detail. This may include the sensitivity of population recovery time to the magnitude and timing of acute (typically lethal) input effects, and the sensitivity of long-term population decline to the magnitude of chronic (typically sublethal) input effects.

Such analyses may provide information on the minimum size of input effects required for a model to show regulatory relevant output effects. E. g., it may be assessed which percent acute mortality would be required in MASTEP to observe a population recovery time longer than a threshold such as eight weeks. Similarly, it may be assessed what extent of decrease in reproduction or what extent of delay in growth would be required in eVole to observe a non-negligible long-term reduction in abundance (specific protection goal for mammals, EFSA 2009b) . Ideally, such analyses should be done across different levels of environmental stress (provided by the environmental scenario). The outcome may be compared with mesocosm or field data and provide essential information on how conservative predictions of a model are in general (see section 4.6 on model validation below).

Additionally, not only the sensitivity of model output to uncertainty in the parameterization, but also to uncertainty in the model design may be assessed; this can be achieved by switching on and off specific processes or by switching between alternative modules for the same process. Though this has been rarely done, a few simple examples can be found in the demonstration of the IBM *Chaoborus* population model and MASTEP in which model output was compared when recolonization or the drift of organisms was switched on and off.

For the $SPEAR_{pesticides}$ approach, a sensitivity analysis may be of limited use, similar to the TKTD models for individual-level effects: The relationship of individual-level vs. community-level effects, that is central in this model, has been calibrated to observational data. However, $SPEAR_{pesticides}$ comes with extensive built-in parameterization regarding the ecological traits used for the classification of the different taxonomic groups (Liess et al. 2008). Therefore, the sensitivity of the approach to the potential misclassification of relevant taxa, to the depth of taxonomic classification, and to the overall number of taxonomic groups in a community may be assessed. Such an analysis would provide essential information on the general robustness of the approach.

4.6 Model Validation

Model validation, i. e. the testing of model predictions with independent experimental or observational data that have not been used for model parameterization, has been identified as a central step of model development in the EFSA Sci. Op. on GMP (EFSA PPR 2014b). However, the value of validation depends on the use of models. TKTD models for individual-level effects may be simply used as a more

precise alternative to the classical dose-response modelling of effects. In this case, validation with independent data would actually assess the repeatability of the test used for calibration, rather than the model itself. However, if a model is used to predict effects from exposure profiles different than those used for calibration, this predictive power may be tested with independent data. Accordingly, the EFSA Sci. Op. on TKDT models (EFSA PPR 2018) recommend specific criteria for validation before a TKTD model may be used for a refined aquatic risk assessment that follows the Tier 2C approach. Because TKTD models have been fully or largely subject to calibration to a specific data set, validation should be performed after each calibration in a new model application (EFSA PPR 2018). TKTD models are also used as toxicity modules in higher level models, i. e. they predict input effects at organism level from a given exposure (see section 4.1). Because exposure in higher-level models may be variable, TKTD models are applied to untested exposure profiles when they are used as toxicity modules in higher-level models. E.g., in an IBM, each individual can experience a different exposure profile, and these profiles will vary from the profile used for the calibration of direct effects on individuals. Accordingly, TKTD models used as toxicity modules in higher-level models should be subjected to the same validation criteria as if they were used in Tier 2C. An example of such a validation attempt can be found in a demonstration of the IBM *Chaoborus* population model coupled with a GUTS module (Dohmen et al. 2016). Otherwise, TKTD models have typically not been validated before they were applied in population modelling studies. Such validation studies are needed; they would be case-specific and should be presented along with a modelling study for the risk assessment of a specific pesticide application.

In a similar way, also other modelling approaches that extrapolate input effects at organism level to untested exposure profiles in higher-level models should be validated after calibration in each case study. This refers e. g. to the scaling of dose-response models for exposure time such as in AQUATOX. It also refers to the application of dose-response models to time-weighted average concentrations (TWA) from exposure profiles that differ from the profile used for calibration, as was done e. g. in eVole and in the application of MASTEP to pyridalyl. These approaches rely on specific assumptions regarding the reciprocity of exposure time and concentration (e. g. Haber's rule, Miller et al. 2000) which do not always hold. Unfortunately, however, it seems that in case studies proposed for risk assessment, toxicity modules of higher-level models have usually not been validated so far (see part 3 of this report).

In contrast, case-specific validation of toxicity modules for individual-level effects in higher-level models may not be relevant when they are only applied to the exposure profile used for calibration: E. g., in most of the evaluated applications of population models, a dose-response module was calibrated to acute Tier 1 toxicity data of a fast-acting and fast-dissipating pesticide. At the day of pesticide application, the module was then executed once to predict mortality based on the current exposure level. This use reflects the probably worst-case assumption that pulse exposure in the field will have the same profile as in the acute toxicity tests (constant exposure for several days). As another example, when MASTEP was applied to deltamethrin, a dose-response module was fitted to the mortality observed in mesocosms after a series of exposure pulses according to the proposed GAP. Effects derived from this toxicity module were executed once, considering that the exposure profile in the mesocosm was representative to the expected exposure profile in the field. In these cases, it may be sufficient to apply similar acceptability criteria as for the general use of dose-response models (e. g. goodness of fit).

Population models themselves have in some, but not all cases been subjected to validation. Because higher-level models come with a default parameterization at least for the physiological part, validation of those models is not fully case-specific but may be considered for a model in general. When population models have been tested, predictions of population dynamics in a control scenario without pesticide effects have been compared with experimental data. Additionally, in some cases the output from individual modules for specific mechanisms has been compared to real world observations. Such testing is helpful to assess the structural integrity of a model, because apparently correct predictions at

the highest level might have resulted by chance from inappropriate mechanisms included. Demonstrating that model predictions do not hold only at the final population level but also for the different modules thus indicates the robustness of a model against unforeseen conditions that may break the model due to highly influential parameters or processes. However, the most relevant model output from a regulatory point of view, i. e. predictions on the population recovery time or the NOAEL, has been addressed only in a single case (demonstration of the IBM *Chaoborus* population model, see section 2.5.2.9).

Therefore, the unsatisfying validation of higher-level models has been identified as a major point for improvement in this report. Validation is typically limited due to the shortage of available ecotoxicological studies for comparison. However, higher-level models may not only be applied to pesticide data, but also to studies on effects from other stressors. The case-specific toxicity modules for input effects at the organism level may need to be adapted, but the ecological mechanisms that drive propagation from input effects to output effects at population and community level, as well as population recovery, will be the same. Therefore, historical data on ecological effects of various toxicants or even other stressors such as desiccation may be used to test model predictions on the vulnerability and recovery of populations and communities in the field.

It may be very informative to establish a general relation of the magnitude of input vs. the magnitude of output effects or vs. population recovery time in a model, ideally across a range of environmental scenarios. This general relationship can then be tested using a number of historical data on various toxicants or even other stressors. If the predicted and the observed increase in population-level effects with individual-level effects match, higher-level models might be safely considered fit for the application in regulatory risk assessment.

In contrast, validation with independent data might be of little use when dealing with the $SPEAR_{pesticides}$ approach. The SPEAR value vs. individual-level effects relationship has been calibrated using a number of independent observational studies, and applying the approach to yet another study might not provide much new information. The situation is comparable to the TKTD models described above, but not case-specific because a number of different studies have been used for calibration.

4.7 Documentation and Access

The assessed models were generally well documented. The models for individual-level effects are publicly available through various software packages and supported with extensive open literature. In some cases, the source code and a manual are also available.

Descriptions of the population models used the ODD protocol for the documentation of IBMs (Grimm et al. 2006, Grimm et al. 2010). Additionally, many models referred to the more comprehensive TRACE protocol (Schmolke et al. 2010, Grimm et al. 2014) for the planning, performing and documentation of quality assurance during the whole process of model development which has been referred to as “evaludation” (Augusiak et al. 2014). However, generally not all aspects in this protocol have been addressed, such as decisions on a specific sub-model for a process for which various alternative sub-models may have been available (see also the discussion on sensitivity analysis above). For ALMaSS, the source code and extensive literature is publicly available and the software may be used after registration to the author’s project. In contrast, the other population models are not freely available and thus no manual has been published (yet). However, the software may be provided to authorities upon request, and for some models an extensive history of documentation and applications exists in the open literature.

The $SPEAR_{pesticides}$ approach may be applied in a similar way as the other models to predict community-level effects (quantified as the SPEAR value) due to a given exposure level (quantified as toxic unit).

For this use of $SPEAR_{pesticides}$, currently no software is available. The free software INDICATE that contains the trait data base for the classification of taxa was designed only for use in the opposite direction of predicting exposure based on observed effects. Since the prediction of SPEAR values from exposure is only based on a simple linear equation, calculations may be done by hand. However, information on the uncertainty of predictions (confidence intervals) so far must be visually assessed from the relationship as presented in Knillmann et al. (2018). AQUATOX for ecosystem-level effects is publicly available as a standalone software for free use and is supported with extensive literature, including a manual, a technical documentation and a collection of sensitivity analyses and applications.

4.8 Model Output Presentation

When a model has passed the steps of development and evaluation outlined in the EFSA Sci. Op. on GMP (2014b) and in the TRACE protocol (Grimm et al. 2014), it may be applied for the regulatory risk assessment of a specific pesticide use. However, as outlined in section 4.2 on model applicability above, there is no agreement on a precise definition of what population recovery time is acceptable and of what kind of long-term effects are negligible, and on how to assess effect size and duration from modelling output. The EFSA Sci. Op. on GMP (2014b) provides little guidance or information on this issue, and also the TRACE protocol is dealing rather with the development but not with the application of models in risk assessment (Grimm et al. 2014). For TKTD models dealing with individual-level effects, the EFSA Sci. Op. on TKTD Modelling EFSA PPR (2018) shows a few examples of model output presentation, but also focuses rather on model development and setup. As a consequence, in the evaluated case studies, model results have been summarized and analysed in various ways. Some studies simply provided raw model outputs in the form of graphs that show the abundance (population size or density) in control and in treatment scenarios over time. All the higher-level models reviewed in detail incorporated variability in several parameters so that the modelling output represented data distributions from repeated model runs instead of a single value for each time point. Therefore, the reviewed studies for risk assessment showed the mean abundance together with a measure of variability such as the range (e. g. application of the IBM *Chaoborus* population model to beta-cyfluthrin) or the 95 % confidence interval (applications of MASTEP to pyridalyl and deltamethrin, application of IDamP to pirimicarb). Such graphs are important to get an overall view on the results, but may not be sufficient because it is difficult to thoroughly assess the magnitude and duration of effects from the provided information. The studies mentioned above focused on the demonstration of population recovery from acute effects. In some cases, additionally an estimate of the mean time for population recovery was presented, sometimes together with a 95 % confidence interval. A number of different criteria were applied to judge when a population had recovered (see also discussion on model applicability in section 4.2 above). This additional information is helpful, though agreed criteria need to be established regarding when a population is considered to be recovered.

A different approach to summarize model results has been presented in the application of eVole to folpet. This study focused on the potential of chronic effects for long-term repercussions on population abundance. Here, the cumulative number of all individuals that died and of those individuals that died due to pesticide exposure was shown over time. This information is useful to assess how population effects accumulate but the graphs are difficult to read due to the high overall number of individuals. Interpretation could be improved by showing the proportion of individuals that died due to the treatment on the cumulative number of all deceived individuals. Additionally, the ratio of the mean treatment vs. control abundance was presented over one year, together with 95 % confidence intervals for the mean control abundance. From this graph it can be assessed whether the mean treatment abundance decreased below a range in which the mean control abundance can be located with 95 % certainty (lower margin of confidence interval). Showing the output in the treatment relative to those in the control is helpful to support an assessment of the effect size.

However, the median may be preferred over the mean to describe the central tendency of a modelling output distribution, because it is less sensitive to outliers. Additionally, a 95 % confidence interval will decrease with an increasing number of replicate model runs (Altman and Bland 2005). The amount of model runs can be set to an arbitrary number (limited only by computational effort) and ranged from only five in the application of MASTEP to pyridalyl to 100 in the application of the IBM *Chaoborus* population model to beta-cyfluthrin. Therefore, a 95 % confidence intervals is of limited use, and instead a measure of variability may be used that does not depend on the number of model runs, such as quantile ranges or the standard deviation (Altman and Bland 2005). Moreover, an issue arises when a potential effect is considered not significant because treatment abundance is within the confidence interval of the control abundance. Confidence intervals and the associated tests assess the risk of assuming an effect while there is actually none (risk of false alarm). If the probability of this type I or α error is below 5 %, the effect is considered significant by convention (Crawley 2005). However, risk assessors of pesticides are interested in the risk of assuming no effect while there is actually one (risk of neglected alarm). The probability of this type II or β error cannot be directly addressed from these statistical tests and the confidence interval, and the error strongly increases the closer results in the treatment are to those in the control.

Therefore, the analysis of significance of a modelled effect should be supplemented with an analysis of relevance based on the effect size. E. g., the median treatment abundance relative to the median control abundance may be plotted over time, together with the interquartile range for both the control and the treatment abundance. The interquartile range covers those 50 % of values from a distribution that are closest to its median. When the median treatment abundance is below the interquartile range of the control abundance, treated populations are on average smaller than 75 % of control populations. This effect size could be considered a threshold for a relevant effect (though different quantile ranges could be considered as well and a general agreement on thresholds for acceptable effect sizes should be sought).

The analysis and interpretation of modelling output may be further improved by the use of pseudo-randomness. This means that random values for initial conditions, parameters and driving variables are drawn prior to a coupled set of control and treatment runs, so that differences between the coupled model runs can be related exclusively to the effects of the pesticide. In the model applications reviewed in detail, the use of pseudo-randomness has been reported only for initial conditions in the application of the *Chaoborus* IBM Population Model to beta-Cyfluthrin.

Finally, the regulatory relevant modelling output such as decrease in abundance or recovery time may be converted to an endpoint that is compatible with other approaches in the ERA of pesticides. E. g., in the application of GUTS to benzovindiflupyr, the concentration that causes 10 % acute mortality (LC10) in the simulations has been calculated as endpoint. The LC10 was divided by the predicted environmental concentrations (PEC) of the pesticide in different FOCUS profiles to obtain margins of safety (factors by which the FOCUS profile must be multiplied to reach 10% mortality at the end of the profile).

However, when population models are used for the refinement of acute risk assessment (see section 4.2.2), the most relevant endpoint is probably the maximum concentration or dose at which the time required for population recovery is still considered acceptable. We consider this the no observable adverse effect concentration or level (NOAEC / NOAEL) for populations (https://www.efsa.europa.eu/en/glossary/noael). In contrast, when population models are used for the refinement of chronic risk assessment, the most relevant endpoint is probably the no observable effect concentration or level (NOEC /NOEL) at and below which no non-negligible effects will occur. Dividing the NO(A)EC / NO(A)EL by the PEC follows the principle of a toxicity-exposure ratio (TER) established in the European framework for risk assessment. This provides a margin of safety (factor by which pesticide exposure could be multiplied before the NO(A)EC /NO(A)EL is exceeded). The margin of safety

could be compared with the uncertainty associated with the predicted NO(A)EC / NO(A)EL (see section 4.9 below) to assess the level of protection in the modelling-based risk assessment. However, no such endpoints and the associated margin of safety have been provided in the reviewed applications of population models for risk assessment.

4.9 Uncertainty Analysis

Uncertainty analysis is considered an important step of model development and application, but is not further described in the EFSA Sci. Op. on GMP (2014b). Uncertainty analysis may refer to the assessment of both uncertainty in model parameters and structure and of how this is propagated to uncertainty associated with model output.

Uncertainty in parameter values can be described by a confidence interval or at least a range of observed values that have been observed for a given parameter. The same applies to uncertainty in other values of model input such as initial conditions and driving variables. In case only one or few observations are available for parameterization (i.e. without sufficient information on the variability in repeated observations), uncertainty should be estimated based on expert judgement. Additionally, it needs to be assessed whether the set of available observations is appropriate (e. g. not biased by a systematic measurement error) and representative for the models' domain of applicability. In case of parameter values that have been obtained from model fitting (calibration), confidence intervals can be estimated during the fitting procedure. Again, it needs to be assessed whether the data set used for calibration is appropriate and representative for the domain of applicability. The uncertainty in basic model parameters that are considered constant across mode applications should be assessed already in the phase of model development (see section 4.1). Many, but not all of such parameters have been reported with an estimate of uncertainty in the model documentations evaluated in part 2 of this report. Specific parameters and input variables are parameterized in the process of a model application and should be then subjected to an uncertainty analysis as well. We found uncertainty estimates and a justification why they are representative for the domain of use only for some of those parameters in the modelling reports evaluated in part 3. In consequence, the analysis of uncertainty in model parameters should be improved and extended to all parameters of a model.

Propagation of uncertainty in model parameters to the joint uncertainty in model output can be estimated using Monte Carlo simulations (Loucks and van Beek 2017). However, this type of analysis is not limited to parametric uncertainty but can and should be applied additionally to assess structural uncertainty in model design. This can be achieved by switching on and off modules for processes whose relevance for a model is unclear, and by switching between alternative modules of a model that describe a given process in different ways. E. g., the process of density regulation in a population model could be simulated through a decrease in survival, reproduction or both with increasing population density. The distribution in model output from a Monte Carlo simulation can be summarized using an error bar or band that shows e. g. the range or the standard deviation around the deterministic or mean probabilistic model output. This error band will be larger than and must be distinguished from an error band that shows variability in output from a probabilistic model.

This way of quantifying and visualizing uncertainty could be a great strength of mechanistic effect models, but has been rarely presented in the reviewed model applications. The error bands presented did typically not present joint uncertainty but only variability in model output. A simple example for the assessment of structural uncertainty can be found in the application of GUTS to benzovindiflupyr (section 3.2); this model has been run with alterative modules for different modes of action that may have caused the observed lethality (individual tolerance vs. stochastic death). Then, the most conservative predictions have been used for risk assessment. As a second example, the IBM Chaoborus Population Model (section 3.6) was applied to a ditch scenario with migration between a treatment and a control population being switch on or off.

Finally, the mean of the predicted model output from an uncertainty analysis can be compared to the deterministic output (or to the mean of a probabilistic model run that includes only variability but no uncertainty). If the mean output from a sensitivity analysis is significantly different from the (mean) output from normal model runs, the chosen parameterization obviously represents a best or worst case and considerably affects the model output. This comparison should not be limited to model output such as population size, but particularly applied also to the predicted output effects of a model in order to assess how conservative predictions of a regulatory model will be.

The uncertainty associated with a predicted regulatory relevant endpoint such as LCx or NO(A)EC / NO(A)EL may then be compared with the margin of safety (factor by which pesticide exposure could be multiplied before non-acceptable effects are predicted, see section 4.8 above).

Uncertainty in the effect predictions of the $SPEAR_{pesticides}$ approach may be assessed from the 95 % confidence intervals for the established SPEAR vs. TU_{max} (effects vs. exposure) relationship from various field observations (see Fig. 3A in Liess et al. 2021). These confidence intervals relate to the joined uncertainty that results from natural variability in the community composition in the field, in the exposure patterns and in other environmental conditions that may affect the sensitivity to pesticides, as well as to the uncertainty from measurement errors. Confidence intervals associated with predictions on effects in the field from $SPEAR_{pesticides}$ thus cover more sources of variability and uncertainty (and are consequently larger) than those presented in the case studies for other models. However, a separation of output variability and uncertainty is not possible for $SPEAR_{pesticides}$. Additionally, uncertainty in model predictions due to uncertainty in the taxa classification and in the decision on a standard reference community or trait composition (that may possibly differ across stream types) may be further assessed.

A sound uncertainty analysis on thoroughly developed and evaluated mechanistic effect models may have a great potential of visualizing and quantitatively assessing uncertainty in the risk assessment of pesticides. This information may not only be used to justify the reduction of assessment factors in a refined risk assessment. Instead, it may help to identify whether the assessment factors currently applied in the conventional approaches of risk assessment are actually protective. These assessment factors are considered to cover all relevant sources of uncertainty which remains largely invisible in a conventional risk assessment procedure, but have been sometimes established decades ago based on potentially outdated information (EFSA 2009b).

However, it should be noted that no uncertainty analysis can provide protection from the risk that modelling results may be totally wrong due to a novel or undetected mode of action. Classical textbook examples include the weakening of egg shells in birds following DDT exposure (Carson 1962) or the unusually high tolerance of the standard test organism *Daphnia magna* to neonicotinoid insecticides (Pisa et al. 2015) that may not be predicted from mechanistic effect models. Therefore, effect models may supplement but should never fully replace physical experiments and field observations for the risk assessment of pesticides.

5 List of Annexes

- Annex 1: Presentations at the Symposium
- Annex 2: Minutes from the Symposium

6 References

Adamczewska-Andrzejewska, K. and L. Nabaglo (1977): Demographic parameters and variations in numbers of the common vole. Acta Theriologica 22: 431-457.

Adams, B. M., H. T. Banks, J. E. Banks and J. D. Stark (2005): Population dynamics models in plant-insect herbivore-pesticide interactions. Mathematical Biosciences 196(1): 39-64.

Addison, R. F. and T. G. Smith (1974): Organochlorine Residue Levels in Arctic Ringed Seals: Variation with Age and Sex. Oikos 25(3): 335-337.

Agatz, A., T. A. Cole, T. G. Preuss, E. Zimmer and C. D. Brown (2013): Feeding inhibition explains effects of imidacloprid on the growth, maturation, reproduction, and survival of *Daphnia magna*. Environmental Science & Technology 47(6): 2909-2917.

Akhurst, L. C. (2005): Folpet, Determination of minimum inhibitory concentrations against selected micro-organisms representative of rabbit gut microflora. Research Laboratory, Huntingdon Life Sciences Limited; Woodley Road, Alconbury Huntingdon, Cambridgeshire, PE28 4HS, England. Report No. MAK 887/052847. Sponsor: Makhteshim Chemical Works Ltd., PO Box 60, Beer-Sheva 84100, Israel. Sponsor Ref. No. R-18665.

Albert, C., R. Ashauer, H. R. Künsch and P. Reichert (2012): Bayesian experimental design for a toxicokinetic-toxicodynamic model. Journal of Statistical Planning and Inference 142(1): 263-275.

Alda Álvarez, O., T. Jager, S. A. L. M. Kooijman and J. E. Kammenga (2005): Responses to stress of *Caenorhabditis elegans* populations with different reproductive strategies. Functional Ecology 19: 656-664.

Alda Álvarez, O., T. Jager, E. Marco Redondo and J. E. Kammenga (2006a): Physiological modes of action of toxic chemicals in the nematode *Acrobeloides nanus*. Environmental Toxicology and Chemistry 25(12): 3230-3237.

Alda Álvarez, O., T. Jager, B. Nuñez Coloa and J. E. Kammenga (2006b): Temporal dynamics of effect concentrations. Environmental Science & Technology 40: 2478-2484.

Altman, D. G. and J. M. Bland (2005): Standard deviations and standard errors. BMJ (Clinical research ed.) 331(7521): 903-903.

Amemiya, T., T. Enomoto, A. G. Rossberg, T. Yamamoto, Y. Inamori and K. Itoh (2007): Stability and dynamical behavior in a lake-model and implications for regime shifts in real lakes. Ecological Modelling 206(1-2): 54-62.

Ananthasubramaniam, B., E. McCauley, K. A. Gust, A. J. Kennedy, E. B. Muller, E. J. Perkins and R. M. Nisbet (2015): Relating suborganismal processes to ecotoxicological and population level endpoints using a bioenergetic model. Ecological Applications 25(6): 1691-1710.

Andersen, T. H., R. Tjørnhøj, L. Wollenberger, T. Slothuus and A. Baun (2006): Acute and chronic effects of pulse exposure of Daphnia magna to dimethoate and pirimicarb. Environmental Toxicology and Chemistry 25(5): 1187-1195.

Ankley, G. T., R. J. Erickson, G. L. Phipps, V. R. Mattson, P. A. Kosian, B. R. Sheedy and J. S. Cox (1995): Effects of light intensity on the phototoxicity of fluoranthene to a benthic macroinvertebrate. Environmental Science & Technology 29(11): 2828-2833.

ANSES (2017): Draft Renewal Assessment Report (DRAR) on Bromoxynil (Bromoxynil-Octanoate 327.5 G). French Agency for Food, Environmental and Occupational Health & Safety, ANSES. R. M. S. F. c.-R. M. S. Germany.

Arnot, J. A. and F. A. Gobas (2004): A food web bioaccumulation model for organic chemicals in aquatic ecosystems. Environ Toxicol Chem 23(10): 2343-2355.

Arzul, G., F. Quiniou and C. Carrie (2006): In vitro test-based comparison of pesticide-induced sensitivity in marine and freshwater phytoplankton. Toxicology Mechanisms and Methods 16(8): 431-437.

Ashauer, R. (2010): Toxicokinetic-toxicodynamic modelling in an individual based context-consequences of parameter variability. Ecological Modelling 221(9): 1325-1328.

Ashauer, R. (2012): Modeling fish survival under dynamic chemical stress. Syngenta; Report No. SYN545192_10325.

Ashauer, R., C. Albert, S. Augustine, N. Cedergreen, S. Charles, V. Ducrot, A. Focks, F. Gabsi, A. Gergs, B. Goussen, T. Jager, N. I. Kramer, A. M. Nyman, V. Poulsen, S. Reichenberger, R. B. Schäfer, P. J. Van den Brink, K. Veltman, S. Vogel, E. I. Zimmer and T. G. Preuss (2016): Modelling survival: exposure pattern, species sensitivity and uncertainty. Scientific Reports 6: 11.

Ashauer, R., A. B. A. Boxall and C. D. Brown (2007a): New ecotoxicological model to simulate survival of aquatic invertebrates after exposure to fluctuating and sequential pulses of pesticides. Environmental Science & Technology 41(4): 1480-1486.

Ashauer, R., A. B. A. Boxall and C. D. Brown (2007b): Simulating toxicity of carbaryl to *Gammarus pulex* after sequential pulsed exposure. Environmental Science & Technology 41(15): 5528-5534.

Ashauer, R. and T. Jager (2018): Physiological modes of action across species and toxicants: the key to predictive ecotoxicology. Environmental Science-Processes & Impacts 20(1): 48-57.

Ashauer, R., I. O'Connor and B. I. Escher (2017): Toxic mixtures in time - the sequence makes the poison. Environmental Science & Technology 51(5): 3084-3092.

Ashauer, R., I. O'Connor, A. Hintermeister and B. I. Escher (2015): Death dilemma and organism recovery in ecotoxicology. Environmental Science & Technology 49(16): 10136-10146.

Ashauer, R., P. Thorbek, J. S. Warinton, J. R. Wheeler and S. Maund (2013): A method to predict and understand fish survival under dynamic chemical stress using standard ecotoxicity data. Environmental Toxicology and Chemistry 32(4): 954-965.

Ashby, K. R. (1967): Studies on the ecology of field mice and voles (Apodemus sylvaticus, Clethrionomys glareolus and Microtus agrestis) in Houghall Wood, Durham. Journal of Zoology 152(4): 389-513.

Augusiak, J., P. J. Van den Brink and V. Grimm (2014): Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. Ecological Modelling 280: 117-128.

Augustine, S., B. Gagnaire, C. Adam-Guillermin and S. A. L. M. Kooijman (2012): Effects of uranium on the metabolism of zebrafish, *Danio rerio*. Aquatic Toxicology 118: 9-26.

Augustine, S., B. Gagnaire, M. Floriani, C. Adam-Guillermin and S. A. L. M. Kooijman (2011): Developmental energetics of zebrafish, *Danio rerio*. Comparative Biochemistry and Physiology, Part A 159(3): 275-283.

Baas, J., S. Augustine, G. M. Marques and J. L. Dorne (2018): Dynamic energy budget models in ecological risk assessment: From principles to applications. Science of the Total Environment 628-629: 249-260.

Baas, J. and S. A. L. M. Kooijman (2015): Sensitivity of animals to chemical compounds links to metabolic rate. Ecotoxicology 24(3): 657-663.

Baas, J., B. P. P. Van Houte, C. A. M. Van Gestel and S. A. L. M. Kooijman (2007): Modelling the effects of binary mixtures on survival in time. Environmental Toxicology and Chemistry 26(6): 1320-1327.

Baláž, I. (2010): The influence of the altitude on somatic characteristics size of common vole (Microtus arvalis) in Slovakia. Ekologia 29(2): 174-181.

Banks, J. E. and J. D. Stark (2004): Aphid response to vegetation diversity and insecticide applications. Agr. Ecosyst. Environ. 103(3): 595-599.

Barber, M. C., L. A. Suarez and R. R. Lassiter (1988): Modeling bioconcentration of nonpolar organic pollutants by fish. Environmental Toxicology and Chemistry 7(7): 545-558.

Barnthouse, L. W. (2004): Quantifying population recovery rates for ecological risk assessment. Environmental Toxicology and Chemistry 23(2): 500-508.

Barsi, A., T. Jager, M. Collinet, L. Lagadic and V. Ducrot (2014): Considerations for test design to accommodate energy-budget models in ecotoxicology: a case study for acetone in the pond snail *Lymnaea stagnalis*. Environmental Toxicology and Chemistry 33(7): 1466-1475.

Bartell, S. M., J. E. Breck, R. H. Gardner and A. L. Brenkert (1986): Individual Parameter Perturbation and Error Analysis of Fish Bioenergetics Models. Canadian Journal of Fisheries and Aquatic Sciences 43(1): 160-168.

Bartell, S. M., R. H. Gardner and R. V. O'Neill (1992). Ecological risk estimation. Lewis Publishers Inc.; Boca Raton, USA.

Bartell, S. M., G. Lefebvre, G. Kaminski, M. Carreau and K. R. Campbell (1999): An ecosystem model for assessing ecological risks in Quebec rivers, lakes, and reservoirs. Ecological Modelling 124(1): 43-67.

Bastiansen, F., J.-D. Ludwigs and I. Barber (2013): Population modeling of small mammals as a higher tier approach in ecological risk assessment: a case study. SETAC conference poster.

Bastiansen, F. and M. Meli (2016): Population modelling for the common vole to assess the potential effects following the application of folpet in vines. RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany. Report No. R1520157. Sponsor: ADAMA Makhteshim Ltd., P.O.Box 60, Industrial Zone Beer-Sheva, 84100, Israel. Sponsor Ref. No. R-36423.

Baveco, J. M., W. H. J. Beltman, G. Fait and P. J. Van den Brink (2012): Estimating recovery times of Asellus aquaticus after pyridalyl exposure by the MASTEP model. Alterra, Wageningen University and Research centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0123. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome,Chuo-ku, Tokyo 104-8260, Japan.

Baveco, J. M. and A. M. De Roos (1996): Assessing the impact of pesticides on lumbricid populations: an individual-based modelling approach. Journal of Applied Ecology 33: 1451-1468.

Baveco, J. M., S. Norman, I. Roessink, N. Galic and P. J. Van den Brink (2014): Comparing population recovery after insecticide exposure for four aquatic invertebrate species using models of different complexity. Environmental Toxicology and Chemistry 33(7): 1517-1528.

Becher, M. A., V. Grimm, P. Thorbek, J. Horn, P. J. Kennedy and J. L. Osborne (2014): BEEHAVE: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure. Journal of Applied Ecology 51(2): 470-482.

Becker, J. M. and M. Liess (2015): Biotic interactions govern genetic adaptation to toxicants. Proceedings of the Royal Society B-Biological Sciences 282(1806): 20150071.

Becker, J. M. and M. Liess (2017): Species Diversity Hinders Adaptation to Toxicants. Environmental Science & Technology 51(17): 10195–10202.

Bedaux, J. J. M. and S. A. L. M. Kooijman (1994): Statistical analysis of bioassays based on hazard modelling. Environmental and Ecological Statistics 1: 303-314.

Bednarska, A. J., P. Edwards, R. Sibly and P. Thorbek (2013): A toxicokinetic model for thiamethoxam in rats: implications for higher-tier risk assessment. Ecotoxicology 22(3): 548-557.

Beketov, M. A., K. Foit, R. B. Schaefer, C. A. Schriever, A. Sacchi, E. Capri, J. Biggs, C. Wells and M. Liess (2009): SPEAR indicates pesticide effects in streams - Comparative use of species- and family-level biomonitoring data. Environmental Pollution 157(6): 1841-1848.

Beketov, M. A. and M. Liess (2008a): Acute and delayed effects of the neonicotinoid insecticide thiacloprid on seven freshwater arthropods. Environ Toxicol Chem 27(2): 461-470.

Beketov, M. A. and M. Liess (2008b): Acute and delayed effects of the neonicotinoid insecticide thiacloprid on seven freshwater arthropods. Environmental Toxicology and Chemistry 27(2): 461-470.

Beketov, M. A. and M. Liess (2008c): An indicator for effects of organic toxicants on lotic invertebrate communities: Independence of confounding environmental factors over an extensive river continuum. Environmental Pollution 156(3): 980-987.

Belden, J. B., R. J. Gilliom and M. J. Lydy (2007): How well can we predict the toxicity of pesticide mixtures to aquatic life? Integr Environ Assess Manag 3(3): 364-372.

Berendonk, T. U. and M. B. Bonsall (2002): The phantom midge and a comparison of metapopulation structures. Ecology 83(1): 116-128.

Berticat, C., O. Duron, D. Heyse and M. Raymond (2004): Insecticide resistance genes confer a predation cost on mosquitoes, *Culex pipiens*. Genetics Research 83(3): 189-196.

Billoir, E., A. da Silva Ferrão-Filho, M. L. Delignette-Muller and S. Charles (2009): DEBtox theory and matrix population models as helpful tools in understanding the interaction between toxic cyanobacteria and zooplankton. Journal of Theoretical Biology 258(3): 380-388.

Billoir, E., H. Delhaye, B. Clément, M. L. Delignette-Muller and S. Charles (2011): Bayesian modelling of daphnid responses to time-varying cadmium exposure in laboratory aquatic microcosms. Ecotoxicology and Environmental Safety 74(4): 693-702.

Billoir, E., M. L. Delignette-Muller, A. R. R. Péry and S. Charles (2008a): A Bayesian approach to analyzing ecotoxicological data. Environmental Science & Technology 42(23): 8978-8984.

Billoir, E., M. L. Delignette-Muller, A. R. R. Péry, O. Geffard and S. Charles (2008b): Statistical cautions when estimating DEBtox parameters. Journal of Theoretical Biology 254(1): 55-64.

Billoir, E., A. R. R. Péry and S. Charles (2007): Integrating the lethal and sublethal effects of toxic compounds into the population dynamics of *Daphnia magna*: a combination of the DEBtox and matrix population models. Ecological Modelling 203(3-4): 204-214.

Bodiguel, X., O. Maury, C. Mellon-Duval, F. Roupsard, A. M. Le Guellec and V. Loizeau (2009): A dynamic and mechanistic model of PCB bioaccumulation in the European hake (*Merluccius merluccius*). Journal of Sea Research 62(2-3): 124-134.

Bohle, H. W. (1995). Spezielle Ökologie - Limnische Systeme. Springer; Marburg, Germany.

Bontje, D., B. W. Kooi, M. Liebig and S. A. L. M. Kooijman (2009a): Modelling long-term ecotoxicological effects on an algal population under dynamic nutrient stress. Water Research 43(13): 3292-3300.

Bontje, D., B. W. Kooi and B. van Hattum (2011): Sublethal toxic effects in a generic aquatic ecosystem. Ecotoxicology and Environmental Safety 74(4): 929-939.

Bontje, D., B. W. Kooi, G. A. K. Van Voorn and S. A. L. M. Kooijman (2009b): Feeding threshold for predators stabilizes predator-prey systems. Mathematical Modelling of Natural Phenomena 4(6): 91-108.

Brett, J. R. and T. D. D. Groves (1979). Physiological energetics. Book chapter in: Fish physiology. VIII Bioenergetics and growth. W. S. Hoar, D. J. Randall and J. R. Brett. Academic Press; New York: 279-352.

Brown, A. R., A. M. Riddle, N. L. Cunningham, T. J. Kedwards, N. Shillabeer and T. H. Hutchinson (2003): Predicting the effects of endocrine disrupting chemicals on fish populations. Human and Ecological Risk Assessment 9(3): 761-788.

Brown, A. R., A. M. Riddle, I. J. Winfield, J. M. Fletcher and J. B. James (2005): Predicting the effects of endocrine disrupting chemicals on healthy and disease impacted populations of perch (perca fluviatilis). Ecological Modelling 189(3-4): 377-395.

Büns, M. and H. T. Ratte (1991): The combined effects of temperature and food consumption on body weight, egg production and developmental time in Chaoborus crystallinus De Geer (Diptera: Chaoboridae). Oecologia 88(4): 470-476.

Burkhard, L. P. (1998): Comparison of two models for predicting bioaccumulation of hydrophobic organic chemicals in a great lakes food web. Environmental Toxicology and Chemistry 17(3): 383-393.

Byers, J. A. and S. J. Castle (2005): Areawide models comparing synchronous versus asynchronous treatments for control of dispersing insect pests. Journal of Economic Entomology 98(6): 1763-1773.

Cabral, J. A., J. C. Marques and S. N. Nielsen (2001): Modeling mosquitofish (*Gambusia holbrooki*) responses to Genapol OXD-080, a non-ionic surfactant, in rice fields. Ecological Engineering 16(4): 537-544.

Calow, P. and R. M. Sibly (1990): A physiological basis of population processes: ecotoxicological implications. Functional Ecology 4(3): 283-288.

Calow, P., R. M. Sibly and V. Forbes (1997): Risk assessment on the basis of simplified life-history scenarios. Environmental Toxicology and Chemistry 16(9): 1983-1989.

Carson, R. (1962). Silent Spring - The classic that launched the environmental movement. Houghton Mifflin Company, Aniversary edition (October 22, 2002).

Caswell, H. (2001). Matrix population models. Sinauer Associates; Sunderland, MA, USA.

Cedergreen, N., N. J. Nørhave, C. Svendsen and D. J. Spurgeon (2016): Variable temperature stress in the nematode *Caenorhabditis elegans* (Maupas) and its implications for sensitivity to an additional chemical stressor. Plos One 11(1): e0140277.

Chandler, G. T., T. L. Cary, A. C. Bejarano, J. Pender and J. L. Ferry (2004): Population consequences of fipronil and degradates to copepods at field concentrations: an integration of life cycle testing with Leslie matrix population modeling. Environmental Science & Technology 38(23): 6407-6414.

Charles, S., M. Ferreol, A. Chaumot and A. R. R. Pery (2004): Food availability effect on population dynamics of the midge *Chironomus riparius*: a Leslie modeling approach. Ecological Modelling 175(3): 217-229.

Chipps, S. R. and D. H. Wahl (2008): Bioenergetics modeling in the 21st century: reviewing new insights and revisiting old constraints. Transactions of the American Fisheries Society 137(1): 298-313.

Chiu, M.-C., L. Hunt and V. H. Resh (2016): Response of macroinvertebrate communities to temporal dynamics of pesticide mixtures: A case study from the Sacramento River watershed, California. Environmental Pollution 219: 89-98.

Chow, T. E., K. F. Gaines, M. E. Hodgson and M. D. Wilson (2005): Habitat and exposure modelling for ecological risk assessment: A case study for the raccoon on the Savannah River Site. Ecological Modelling 189(1): 151-167.

Clough, J. S. (2014): AQUATOX (Release 3.1 plus) - User's Manual. United States Environmental Protection Agency (USEPA). Office of Water (4305). Office of Science and Technology; Washington DC 20460 , USA. Report No. EPA-820-R-14-005.

Cockroft, R. (2017): Beta-cyfluthrin (Bulldock 25EC) – Effects on Chaoborus first instar larvae. Agrochemex; Report No. ENV-16-030.

Collins, T. F. X., J. J. Welsh, T. N. Black, K. E. Whitby and M. W. O'Donnell (1987): Potential reversibility of skeletal effects in rats exposed in utero to caffeine. Food and Chemical Toxicology 25(9): 647-662.

Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union, L 93, p. 1. 03.04.2013.

Commission Regulation (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union, L 93, p. 85. 03.04.2013.

Commission Regulation (EU) No 546/2011 of 10 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards uniform principles for evaluation and authorisation of plant protection products. Official Journal of the European Union, L 155, p. 27. 11.06.2011.

Corbet, G. B. and S. Harris (1996). The Handbook of British Mammals. WileyBlackwell; Oxford, UK.

Council Directive 97/57/EC of 22 September 1997 establishing Annex VI to Directive 91/414/EEC concerning the placing of plant protection products on the market. Official Journal of the European Communities, L 265, p. 87.

Crawley, M. J. (2005). Statistics - An Introduction using R. John Wiley & Sons Ltd.; The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.

Crocker, D. R., A. D. M. Hart, J. Gurney and C. McCoy (2002): Methods for estimating daily food intake of wild birds and mammals, Appendix I. Central Science Laboratory; York, UK.

CTGB (2012): Draft Assessment Report (DAR) on Pyridalyl. Dutch Board for the Authorisation of Plant Protection Products and Biocides, Ctgb. Rapporteur Member State The Netherlands.

Dalkvist, T., R. M. Sibly and C. J. Topping (2013): Landscape structure mediates the effects of a stressor on field vole populations. Landscape Ecology 28(10): 1961-1974.

Dalkvist, T., C. J. Topping and V. E. Forbes (2009): Population-level impacts of pesticide-induced chronic effects on individuals depend more on ecology than toxicology. Ecotoxicology and Environmental Safety 72(6): 1663-1672.

De Bruijn, J. and J. Hermens (1991): Uptake and elimination kinetics of organophosphorous pesticides in the guppy (*Poecilia reticulata*): correlations with the octanol/water partition coefficient. Environmental Toxicology and Chemistry 10: 791-804.

De Laender, F., K. A. De Schamphelaere, C. R. Janssen and P. A. Vanrolleghem (2007): An ecosystem modelling approach for deriving water quality criteria. Water Sci Technol 56(6): 19-27.

de los Santos, C. B., T. Neuparth, T. Torres, I. Martins, I. Cunha, D. Sheahan, T. McGowan and M. M. Santos (2015): Ecological modelling and toxicity data coupled to assess population recovery of marine amphipod Gammarus locusta: Application to disturbance by chronic exposure to aniline. Aquatic Toxicology 163: 60-70.

De Roos, A. M., O. Diekmann and J. A. J. Metz (1992): Studying the dynamics of structured population models: a versatile technique and its application to *Daphnia*. The American Naturalist 139(1): 123-147.

Desneux, N., A. Decourtye and J.-M. Delpuech (2006): The Sublethal Effects of Pesticides on Beneficial Arthropods. Annual Review of Entomology 52(1): 81-106.

Diepens, N. J., W. H. J. Beltman, A. A. Koelmans and P. J. Van den Brink (2016): Dynamics and recovery of a sediment-exposed *Chironomus riparius* population: a modelling approach. Environmental Pollution 213: 741-750.

Dohmen, G. P., T. G. Preuss, M. Hamer, N. Galic, T. Strauss, P. J. Van den Brink, F. De Laender and S. Bopp (2016): Population-level effects and recovery of aquatic invertebrates after multiple applications of an insecticide. Integrated Environmental Assessment and Management 12(1): 67-81.

Driever, S. M., E. H. van Nes and R. M. M. Roijackers (2005): Growth limitation of *Lemna minor* due to high plant density. Aquatic Botany 81(3): 245-251.

Ducrot, V., R. Ashauer, A. J. Bednarska, S. Hinarejos, P. Thorbek and G. Weyman (2015): Using toxicokinetic-toxicodynamic modeling as an acute risk assessment refinement approach in vertebrate ecological risk assessment. Integrated Environmental Assessment and Management 12(1): 32-45.

Ducrot, V., A. R. R. Péry, R. Mons, H. Quéau, S. Charles and J. Garric (2007): Dynamic energy budget as a basis to model population-level effects of zinc-spiked sediments in the gastropod *Valvata piscinalis*. Environmental Toxicology and Chemistry 26(8): 1774-1783.

Duron, O., P. Labbe, C. Berticat, F. Rousset, S. Guillot, M. Raymond and M. Weill (2006): High *Wolbachia* density correlates with cost of infection for insecticide resistant *Culex pipiens* mosquitoes. Evolution 60(2): 303-314.

ECHA (2008). Guidance on information requirements and chemical safety assessment. Chapter R.10: Characterisation of dose [concentration]-response for environment. European Chemicals Agency; Helsinki, Finland.

Effler, S. W. (1996). Limnological and Engineering Analysis of a polluted Urban Lake. Springer Verlag; New York, USA.

EFSA (2005): Conclusion regarding the peer review of the pesticide risk assessment of the active substance pirimicarb. EFSA Scientific Report 43: 1-76.

EFSA (2009a): Conclusion regarding the peer review of the pesticide risk assessment of the active substance folpet. EFSA Scientific Report 297: 1-80.

EFSA (2009b): Guidance of EFSA: Risk Assessment for Birds and Mammals. EFSA Journal 7(12): 1438.

EFSA PPR (2010): Scientific Opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the Guidance Documents on Aquatic and Terrestrial Ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). EFSA Journal 8(10): 1821.

EFSA PPR (2013): Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 11(7): 3290.

EFSA PPR (2014a): Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants. EFSA Journal 12(7): 3800.

EFSA PPR (2014b): Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA Journal 12(3): 92.

EFSA PPR (2015a): Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 13(2): 3996.

EFSA PPR (2015b): Statement on the suitability of the BEEHAVE model for its potential use in a regulatory context and for the risk assessment of multiple stressors in honeybees at the landscape level. EFSA Journal 13(91).

EFSA PPR (2017): Scientific Opinion addressing the state of the science on risk assessment of plant protection products for in-soil organisms. EFSA Journal 15(2): 4690.

EFSA PPR (2018): Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms. EFSA Journal 16(8): e05377.

Eichinger, M., V. Loizeau, F. Roupsard, A. M. Le Guellec and C. Bacher (2010): Modelling growth and bioaccumulation of Polychlorinated biphenyls in common sole (*Solea solea*). Journal of Sea Research 64(3): 373-385.

El Husseiny, I., H. Elbrense, T. Roeder and S. El Kholy (2018): Hormonal modulation of cannibalistic behaviors in mosquito (Culex pipiens) larvae. Journal of Insect Physiology 109: 144-148.

Elton, C., E. B. Ford, J. R. Baker and A. D. Gardner (1931): 36. The Health and Parasites of a Wild Mouse Population. Proceedings of the Zoological Society of London 101(3): 657-721.

Englund, G. and P. A. Hambäck (2004): Scale-dependence of movement rates in stream invertebrates. Oikos 105(1): 31-40.

Escher, B. I., R. Ashauer, S. Dyer, J. L. M. Hermens, H. L. Lee, H. A. Leslie, P. Mayer, J. P. Meador and M. S. J. Warne (2011): Crucial role of mechanisms and modes of toxic action for understanding tissue residue toxicity and internal effect concentrations of organic chemicals. Integrated Environmental Assessment and Management 7(1): 28-49.

Escher, B. I. and J. L. M. Hermens (2002): Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. Environmental Science & Technology 36(20): 4201-4217.

EU Pesticides Database (2018). http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=homepage&language=EN

European Union (2019). http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.detail&language=DE&selectedID=102222.01.2019.

Eurostat (2020). https://ec.europa.eu/eurostat/de/web/products-eurostat-news/-/DDN-20200603-116.12.2020.

Eynaud, Y., R. M. Nisbet and E. B. Muller (2011): Impact of excess and harmful radiation on energy budgets in scleractinian corals. Ecological Modelling 222(7): 1315-1322.

Fabre, F., J. S. Pierre, C. A. Dedryver and M. Plantegenest (2006): Barley yellow dwarf disease risk assessment based on Bayesian modelling of aphid population dynamics. Ecological Modelling 193(3-4): 457-466.

Ferreira, J. G. (1995): ECOWIN - an object-oriented ecological model for aquatic ecosystems. Ecological Modelling 79: 21-34.

Ferreira, J. G., P. Duarte and B. Ball (1998): Trophic capacity of Carlingford Lough for oyster culture – analysis by ecological modelling. Aquatic Ecology 31: 361-378.

Fiechter, J., D. D. Huff, B. T. Martin, D. W. Jackson, C. A. Edwards, K. A. Rose, E. N. Curchitser, K. S. Hedstrom, S. T. Lindley and B. K. Wells (2015): Environmental conditions impacting juvenile Chinook salmon growth off central California: An ecosystem model analysis. Geophysical Research Letters 42(8): 2910-2917.

Flowerdew, J. R. and F. Tattersall (2008). Rodents: order rodentia – wood mouse Apodemus sylvaticus. Book chapter in: Mammals of the British Isles: Handbook. S. Harris and D. W. Yalden. The Mammal Society; Southampton, UK.

Focks, A., D. Belgers, M. C. Boerwinkel, L. Buijse, I. Roessink and P. J. Van den Brink (2018): Calibration and validation of toxicokinetic-toxicodynamic models for three neonicotinoids and some aquatic macroinvertebrates. Ecotoxicology.

Focks, A., R. Luttik, M. Zorn, T. Brock, E. Roex, T. Van der Linden and P. J. Van den Brink (2014a): A simulation study on effects of exposure to a combination of pesticides used in an orchard and tuber crop on the recovery time of a vulnerable aquatic invertebrate. Environmental Toxicology and Chemistry 33(7): 1489-1498.

Focks, A., M. ter Horst, E. van den Berg, H. Baveco and P. J. van den Brink (2014b): Integrating chemical fate and population-level effect models for pesticides at landscape scale: New options for risk assessment. Ecological Modelling 280: 102-116.

FOCUS (2001): FOCUS Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC. Report of the FOCUS Working Group on Surface Water Scenarios. Report No. SANCO/4802/2001.

Foit, K., O. Kaske and M. Liess (2012): Competition increases toxicant sensitivity and delays the recovery of two interacting populations. Aquatic Toxicology 106: 25-31.

Franco, A. and S. Trapp (2010): A multimedia activity model for ionizable compounds: Validation study with 2,4-dichlorophenoxyacetic acid, aniline, and trimethoprim. Environmental Toxicology and Chemistry 29(4): 789-799.

Fujisawa, T., K. Ichise, M. Fukushima, T. Katagi and Y. Takimoto (2002): Mathematical model of the uptake of non-ionized pesticides by edible root of root crops. Journal of Pesticide Science 27(3): 242-248.

Gabsi, F., D. S. Glazier, M. Hammers-Wirtz, H. T. Ratte and T. G. Preuss (2014a): How do interactive maternal traits and environmental factors determine offspring size in Daphnia magna? Annales De Limnologie-International Journal of Limnology 50(1): 9-18.

Gabsi, F., M. Hammers-Wirtz, V. Grimm, A. Schaffer and T. G. Preuss (2014b): Coupling different mechanistic effect models for capturing individual- and population-level effects of chemicals: Lessons from a case where standard risk assessment failed. Ecological Modelling 280: 18-29.

Gabsi, F., M. Hammers-Wirtz, V. Grimm, A. Schäffer and T. G. Preuss (2014c): Coupling different mechanistic effect models for capturing individual- and population-level effects of chemicals: Lessons from a case where standard risk assessment failed. Ecological Modelling 280: 18-29.

Gabsi, F. and T. G. Preuss (2014): Modelling the impact of the environmental scenario on population recovery from chemical stress exposure: A case study using Daphnia magna. Aquatic Toxicology 156: 221-229.

Gabsi, F., A. Schaffer and T. G. Preuss (2014d): Predictiing the sensitivity of poulations from individual exposure to chemical: The role of ecological interactions. Environmental Toxicology and Chemistry 33(7): 1449-1457.

Galic, N., R. Ashauer, H. Baveco, A. M. Nyman, A. Barsi, P. Thorbek, E. Bruns and P. J. Van den Brink (2014): Modeling the contribution of toxicokinetic and toxicodynamic processes to the recovery of *Gammarus pulex* populations after exposure to pesticides. Environmental Toxicology and Chemistry 33(7): 1476-1488.

Galic, N., H. Baveco, G. M. Hengeveld, P. Thorbek, E. Bruns and P. J. van den Brink (2012): Simulating population recovery of an aquatic isopod: Effects of timing of stress and landscape structure. Environmental Pollution 163: 91-99.

Galic, N., U. Hommen, H. Baveco and P. J. Van den Brink (2010): Potential Application of Population Models in the European Ecological Risk Assessment of Chemicals II: Review of Models and Their Potential to Address Environmental Protection Aims. Integrated Environmental Assessment and Management 6(3): 338-360.

Ganatra, A. A., F. J. Kandie, U. Fillinger, F. McOdimba, B. Torto, W. Brack, M. Liess, H. Hollert and J. M. Becker (2021): Calibration of the SPEARpesticides bioindicator for cost-effective pesticide monitoring in East African streams. Environmental Sciences Europe 33(1): 58.

Ganzelmeier, H., D. Rautmann, R. Spangenberg, M. Streloke, M. Herrmann, H.-J. Wenzelburger and H.-F. Walter (1995): Studies on the spray drift of plant protection products - Results of a test program carried out throughout the Federal Republic of Germany. Mitteilungen aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft 305.

Gergs, A., A. Zenker, V. Grimm and T. G. Preuss (2013a): Chemical and natural stressors combined: from cryptic effects to population extinction. Scientific Reports 3: 2036.

Gergs, A., A. Zenker, V. Grimm and T. G. Preuss (2013b): Chemical and natural stressors combined: from cryptic effects to population extinction. Scientific Reports 3.

Gerner, N. V., M. Koné, M. S. Ross, A. Pereira, A. C. Ulrich, J. W. Martin and M. Liess (2017): Stream invertebrate community structure at Canadian oil sands development is linked to concentration of bitumen-derived contaminants. Science of The Total Environment 575: 1005-1013.

Gerritsen, A., N. van der Hoeven and A. Pielaat (1998): The acute toxicity of selected alkylphenols to young and adult *Daphnia magna*. Ecotoxicology and Environmental Safety 39(3): 227-232.

Gledhill, M. and R. W. Van Kirk (2011): Modeling effects of toxin exposure in fish on long-term population size, with an application to selenium toxicity in bluegill (Lepomis macrochirus). Ecological Modelling 222(19): 3587-3597.

Gobas, F. A. and J. A. Arnot (2010): Food web bioaccumulation model for polychlorinated biphenyls in San Francisco Bay, California, USA. Environ Toxicol Chem 29(6): 1385-1395.

Gobas, F. A. P. C. (1993): A model for predicting the bioaccumulation of hydrophobic organic chemicals in aquatic food-webs: application to Lake Ontario. Ecological Modelling 69(1): 1-17.

Gobas, F. A. P. C., A. Opperhuizen and O. Hutzinger (1986): Bioconcentration of hydrophobic chemicals in fish: relationship with membrane permeation. Environmental Toxicology and Chemistry 5(7): 637-646.

Goussen, B., R. Beaudouin, M. Dutilleul, A. Buisset-Goussen, J. M. Bonzom and A. R. R. Péry (2015): Energy-based modelling to assess effects of chemicals on *Caenorhabditis elegans*: a case study on uranium. Chemosphere 120: 507-514.

Grimm, V., J. Augusiak, A. Focks, B. M. Frank, F. Gabsi, A. S. A. Johnston, C. Liu, B. T. Martin, M. Meli, V. Radchuk, P. Thorbek and S. F. Railsback (2014): Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling 280: 129-139.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jorgensen, W. M. Mooij, B. Mueller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rueger, E. Strand, S. Souissi, R. A. Stillman, R. Vabo, U. Visser and D. L. DeAngelis (2006): A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198(1-2): 115-126.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske and S. F. Railsback (2010): The ODD protocol: A review and first update. Ecological Modelling 221(23): 2760-2768.

Grimm, V. and S. F. Railsback (2012): Pattern-oriented modelling: a multi-scope for predictive systems ecology. Philosophical Transactions of the Royal Society B: Biological Sciences 367(1586): 298-310.

Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H. H. Thulke, J. Weiner, T. Wiegand and D. L. DeAngelis (2005): Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. Science 310(5750): 987-991.

Groeneveld, J., K. Johst, S. Kawaguchi, B. Meyer, M. Teschke and V. Grimm (2015): How biological clocks and changing environmental conditions determine local population growth and species distribution in Antarctic krill (*Euphausia superba*): a conceptual model. Ecological Modelling 303: 78-86.

H. Tattersall, F., D. W. Macdonald, J. Hart and B. W. Manley (2004): Balanced dispersal or source–sink–do both models describe wood mice in farmed landscapes? Oikos 106(3): 536-550.

Hallam, T. G., G. A. Canziani and R. R. Lassiter (1993): Sublethal narcosis and population persistence: a modeling study on growth effects. Environmental Toxicology and Chemistry 12(5): 947-954.

Hallam, T. G., R. R. Lassiter and S. M. Henson (2000): Modeling fish population dynamics. Nonlinear Analysis-Theory Methods & Applications 40(1-8): 227-250.

Hallam, T. G., R. R. Lassiter, J. Li and W. McKinney (1990a): Toxicant-induced mortality in models of *Daphnia* populations. Environmental Toxicology and Chemistry 9(5): 597-621.

Hallam, T. G., R. R. Lassiter, J. Li and L. A. Suarez (1990b): Modeling individuals employing an integrated energy response: application to *Daphnia*. Ecology 71(3): 938-954.

Hamda, N. T. (2013). Mechanistic models to explore combined effects of toxic chemicals and natural stressing factors: case study on springtails. PhD thesis Jagiellonian University, Krakow, Poland/VU University, Amsterdam, the Netherlands. http://hdl.handle.net/1871/50121.

Hamda, N. T., V. E. Forbes, J. D. Stark and R. Laskowski (2014): Stochastic density-dependent matrix model for extrapolating individual-level effects of chemicals to the population: Case study on effects of Cd on *Folsomia candida*. Ecological Modelling 280: 53-64.

Hamilton, P. T., J. M. L. Richardson and B. R. Anholt (2012): Daphnia in tadpole mesocosms: trophic links and interactions with Batrachochytrium dendrobatidis. Freshwater Biology 57(4): 676-683.

Hammers-Wirtz, M. and H. T. Ratte (2000): Offspring fitness in *Daphnia*: is the *Daphnia* reproduction test appropriate for extrapolating effects on the population level? Environmental Toxicology and Chemistry 19(7): 1856-1866.

Hanegraaf, P. P. F. and E. B. Muller (2001): The dynamics of the macromolecular composition of biomass. Journal of Theoretical Biology 212(2): 237-251.

Hanson, N. (2011): Population modeling of the fish species roach (*Rutilus rutilus*) to investigate how latitudinal variations in life history traits affect the sensitivity to anthropogenic stress. Environmental Modeling & Assessment 16(4): 401-411.

Hanson, N. and J. D. Stark (2011): A comparison of simple and complex population models to reduce uncertainty in ecological risk assessments of chemicals: example with three species of Daphnia. Ecotoxicology 20(6): 1268-1276.

Hansson, L. (1977): Spatial Dynamics of Field Voles Microtus agrestis in Heterogeneous Landscapes. Oikos 29(3): 539-544.

Hardstone, M. C., B. P. Lazzaro and J. G. Scott (2009): The effect of three environmental conditions on the fitness of cytochrome P450 monooxygenase-mediated permethrin resistance in *Culex pipiens quinquefasciatus*. BMC Evol Biol 9: 42.

Havertz, U. (1988). Einfluss der Temperatur auf die Populationsdynamik von Chaoborus crystallinus in Freiland-Mesokosmen. Diploma Thesis, RWTH Aachen University.

Heimbach, F. (2000): Biological Effects and Fate of Cyfluthrin EC 050 in Outdoor Microcosm Ponds. Bayer AG, Crop Protection, Institue for Environmental Biology; Document No. HBF/Bt 02.

Heine, S. (2014). Development and specification of a toxicokinetic and toxicodynamic growth model of *Myriophyllum spicatum* for use in risk assessment. PhD Thesis, RWTH Aachen UniversityAachen.

Heine, S., W. Schmitt, G. Görlitz, A. Schäffer and T. G. Preuss (2014): Effects of light and temperature fluctuations on the growth of Myriophyllum spicatum in toxicity tests-a model-based analysis. Environmental Science and Pollution Research 21(16): 9644-9654.

Heine, S., W. Schmitt, A. Schaeffer, G. Goerlitz, H. Buresova, G. Arts and T. G. Preuss (2015): Mechanistic modelling of toxicokinetic processes within Myriophyllum spicatum. Chemosphere 120: 292-298.

Hendriks, A. J. and E. L. Enserink (1996): Modelling response of single-species populations to microcontaminants as a function of species size with examples for waterfleas (*Daphnia magna*) and cormorants (*Phalacrocorax carbo*). Ecological Modelling 88(1-3): 247-262.

Hendriks, A. J., A. van der Linde, G. Cornelissen and D. T. H. M. Sijm (2001): The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight. Environmental Toxicology and Chemistry 20(7): 1399-1420.

Hess, M., G. Barralis, H. Bleiholder, L. Buhr, T. H. Eggers, H. Hack and R. Stauss (1997): Use of the extended BBCH scale—general for the descriptions of the growth stages of mono; and dicotyledonous weed species. Weed Research 37(6): 433-441.

Hickey, G. L. and P. S. Craig (2012): Competing statistical methods for the fitting of normal species sensitivity distributions: recommendations for practitioners. Risk Anal 32(7): 1232-1243.

Hickie, B. E., M. C. S. Kingsley, P. V. Hodson, D. C. G. Muir, P. Béland and D. Mackay (2000): A modelling-based perspective on the past, present, and future polychlorinated biphenyl contamination of the St. Lawrence beluga whale (Delphinapterus leucas) population. Canadian Journal of Fisheries and Aquatic Sciences 57(S1): 101-112.

Hickie, B. E., D. C. G. Muir, R. F. Addison and P. F. Hoekstra (2005): Development and application of bioaccumulation models to assess persistent organic pollutant temporal trends in arctic ringed seal (Phoca hispida) populations. Science of the Total Environment 351: 413-426.

Hommen, U. and F. Heimbach (2000): Evaluation of an Outdoor Microcosm Study on Cyfluthrin (Report No. HBF/Bt 02 of March 15, 2000) for an Aquatic Risk Assessment. Bayer AG, Crop Protection, Environmental Biology; Report No. HBF/Bt 02a.

Hommen, U., W. Schmitt, S. Heine, T. C. M. Brock, S. Duquesne, P. Manson, G. Meregalli, H. Ochoa-Acuna, P. van Vliet and G. Arts (2016): How TK-TD and population models for aquatic macrophytes could support the risk assessment for plant protection products. Integrated Environmental Assessment and Management 12(1): 82-95.

Horn, J., M. A. Becher, P. J. Kennedy, J. L. Osborne and V. Grimm (2016): Multiple stressors: using the honeybee model BEEHAVE to explore how spatial and temporal forage stress affects colony resilience. Oikos 125(7): 1001-1016.

Hunt, L., C. Bonetto, N. Marrochi, A. Scalise, S. Fanelli, M. Liess, M. J. Lydy, M. C. Chiu and V. H. Resh (2017): Species at Risk (SPEAR) index indicates effects of insecticides on stream invertebrate communities in soy production regions of the Argentine Pampas. Science of The Total Environment 580: 699-709.

Ibrahim, L., F. P. Podsedensek, A. Van Den Ende, T. Wittwer, F. Bastiansen, O. Jakoby and J. D. Ludwigs (2019): A common issue - Wildlife risk assessment supported by population modelling: a case study on potential herbicide effects on common vole populations. SETAC conference poster.

ISO (2006). Water quality - Guidance on statistical interpretation of ecotoxicity data. International Organization for Standardization (ISO); Geneve, Switzerland.

Jacob, J. (2000). Populationsökologische Untersuchungen an Kleinnagern auf unterschiedlich bewirtschafteten Flächen der Unstrut-Aue. PhD Thesis, Friedrich-Schiller University Jena, Institut für Ökologie und Evolution, Professur für Ökologie; Dornburger Str. 159, 07743 Jena, Germany.

Jacquez, J. A., R. Bellman and R. Kalaba (1960): Some mathematical aspects of chemotherapy II: the distribution of a drug in the body. Bulletin of Mathematical Biophysics 22: 309-322.

Jager, T. (2011): Some Good Reasons to Ban ECx and Related Concepts in Ecotoxicology. Environmental Science & Technology 45(19): 8180-8181.

Jager, T. (2014): Reconsidering sufficient and optimal test design in acute toxicity testing. Ecotoxicology 23(1): 38-44.

Jager, T. (2018). DEBkiss. A simple framework for animal energy budgets. Leanpub, https://leanpub.com/debkiss_book; Version 2.0, 15 November 2018.

Jager, T. (2019). Making sense of chemical stress. Applications of Dynamic Energy Budget theory in ecotoxicology and stress ecology. Leanpub, https://leanpub.com/debtox_book; Version 2.0, 9 May 2019.

Jager, T. (2020): Revisiting simplified DEBtox models for analysing ecotoxicity data. Ecological Modelling 416: 108904.

Jager, T., C. Albert, T. G. Preuss and R. Ashauer (2011): General Unified Threshold Model of Survival - a Toxicokinetic-Toxicodynamic Framework for Ecotoxicology. Environmental Science & Technology 45(7): 2529-2540.

Jager, T., O. Alda Álvarez, J. E. Kammenga and S. A. L. M. Kooijman (2005): Modelling nematode life cycles using dynamic energy budgets. Functional Ecology 19: 136-144.

Jager, T., D. Altin, C. Miljeteig and B. H. Hansen (2016a): Stage-dependent and sex-dependent sensitivity to water-soluble fractions of fresh and weathered oil in the marine copepod Calanus finmarchicus. Environmental Toxicology and Chemistry 35(3): 728-735.

Jager, T., F. A. Antón Sánchez, B. Muijs, E. G. Van der Velde and L. Posthuma (2000): Toxicokinetics of polycyclic aromatic hydrocarbons in *Eisenia andrei* (Oligochaeta) using spiked soil. Environmental Toxicology and Chemistry 19: 953-961.

Jager, T. and R. Ashauer (2018a): How to evaluate the quality of toxicokinetic-toxicodynamic models in the context of environmental risk assessment. Integrated Environmental Assessment and Management 14(5): 604-614.

Jager, T. and R. Ashauer (2018b): Modelling survival under chemical stress. A comprehensive guide to the GUTS framework. York, UK, Toxicodynamics Ltd.

Jager, T., A. Barsi, N. T. Hamda, B. T. Martin, E. I. Zimmer and V. Ducrot (2014a): Dynamic energy budgets in population ecotoxicology: applications and outlook. Ecological Modelling 280: 140-147.

Jager, T., T. Crommentuijn, C. A. M. Van Gestel and S. A. L. M. Kooijman (2004): Simultaneous modeling of multiple endpoints in life-cycle toxicity tests. Environmental Science & Technology 38: 2894-2900.

Jager, T., T. Crommentuijn, C. A. M. Van Gestel and S. A. L. M. Kooijman (2007): Chronic exposure to chlorpyrifos reveals two modes of action in the springtail *Folsomia candida*. Environmental Pollution 145: 452-458.

Jager, T., R. H. L. J. Fleuren, E. A. Hogendoorn and G. De Korte (2003): Elucidating the routes of exposure for organic chemicals in the earthworm, *Eisenia andrei* (Oligochaeta). Environmental Science & Technology 37: 3399-3404.

Jager, T., E. M. Gudmundsdóttir and N. Cedergreen (2014b): Dynamic modeling of sub-lethal mixture toxicity in the nematode *Caenorhabditis elegans*. Environmental Science & Technology 48(12): 7026-7033.

Jager, T., E. H. W. Heugens and S. A. L. M. Kooijman (2006): Making sense of ecotoxicological test results: towards application of process-based models. Ecotoxicology 15: 305-314.

Jager, T. and C. Klok (2010): Extrapolating toxic effects on individuals to the population level: the role of dynamic energy budgets. Philosophical Transactions of the Royal Society B-Biological Sciences 365: 3531-3540.

Jager, T. and S. A. L. M. Kooijman (2005): Modeling receptor kinetics in the analysis of survival data for organophosphorus pesticides. Environmental Science & Technology 39: 8307-8314.

Jager, T. and S. A. L. M. Kooijman (2009): A biology-based approach for quantitative structure-activity relationships (QSARs) in ecotoxicity. Ecotoxicology 18: 187-196.

Jager, T., B. T. Martin and E. I. Zimmer (2013): DEBkiss or the quest for the simplest generic model of animal life history. Journal of Theoretical Biology 328: 9-18.

Jager, T., R. Nepstad, B. H. Hansen and J. Farkas (2018): Simple energy-budget model for yolk-feeding stages of Atlantic cod (Gadus morhua). Ecological Modelling 385: 213-219.

Jager, T., I. B. Øverjordet, R. Nepstad and B. H. Hansen (2017a): Dynamic links between lipid storage, toxicokinetics and mortality in a marine copepod exposed to dimethylnaphthalene. Environmental Science & Technology 51(13): 7707-7713.

Jager, T. and E. Ravagnan (2015): Parameterising a generic model for the dynamic energy budget of Antarctic krill, *Euphausia superba*. Marine Ecology Progress Series 519: 115-128.

Jager, T. and E. Ravagnan (2016): Modelling growth of northern krill (*Meganyctiphanes norvegica*) using an energy-budget approach. Ecological Modelling 325: 28-34.

Jager, T., E. Ravagnan and S. Dupont (2016b): Near-future ocean acidification impacts maintenance costs in sea-urchin larvae: Identification of stress factors and tipping points using a DEB modelling approach. Journal of Experimental Marine Biology and Ecology 474: 11-17.

Jager, T., I. Salaberria, D. Altin, T. Nordtug and B. H. Hansen (2017b): Modelling the dynamics of growth, development and lipid storage in the marine copepod *Calanus finmarchicus*. Marine Biology 164(1): 15.

Jager, T. and H. Selck (2011): Interpreting toxicity data in a DEB framework: a case study for nonylphenol in the marine polychaete *Capitella teleta*. Journal of Sea Research 66: 456-462.

Jager, T., T. Vandenbrouck, J. Baas, W. M. De Coen and S. A. L. M. Kooijman (2010): A biology-based approach for mixture toxicity of multiple endpoints over the life cycle. Ecotoxicology 19: 351-361.

Jager, T. and E. I. Zimmer (2012): Simplified Dynamic Energy Budget model for analysing ecotoxicity data. Ecological Modelling 225: 74-81.

Janssens, L. and R. Stoks (2012): How does a pesticide pulse increase vulnerability to predation? Combined effects on behavioral antipredator traits and escape swimming. Aquatic Toxicology 110: 91-98.

Janssens, L. and R. Stoks (2013): Fitness effects of chlorpyrifos in the damselfly Enallagma cyathigerum strongly depend upon temperature and food level and can bridge metamorphosis. Plos One 8(6): e68107.

Janz, P., L. Weltje, K. P. Ebke and U. Dawo (2016): Temporal population dynamics of the phantom midge Chaoborus crystallinus and its influence on the zooplankton community. Hydrobiologia 770(1): 273-287.

Jenkins, W. R. (2014): Beta-cyfluthrin (Bulldock 25EC): Toxicity to Asellus, Crangonyx, Chaoborus and Cloeon, in Outdoor Microcosms. Huntingdon Life Sciences; Report No. JDV0118.

Jepsen, J. U., C. J. Topping, P. Odderskaer and P. N. Andersen (2005): Evaluating consequences of land-use strategies on wildlife populations using multiple-species predictive scenarios. Agriculture Ecosystems & Environment 105(4): 581-594.

Johnston, A. S. A., M. E. Hodson, P. Thorbek, T. Alvarez and R. M. Sibly (2014a): An energy budget agent-based model of earthworm populations and its application to study the effects of pesticides. Ecological Modelling 280: 5-17.

Johnston, A. S. A., M. Holmstrup, M. E. Hodson, P. Thorbek, T. Alvarez and R. M. Sibly (2014b): Earthworm distribution and abundance predicted by a process-based model. Applied Soil Ecology 84: 112-123.

Johnston, A. S. A., R. M. Sibly, M. E. Hodson, T. Alvarez and P. Thorbek (2015): Effects of agricultural management practices on earthworm populations and crop yield: validation and application of a mechanistic modelling approach. Journal of Applied Ecology 52(5): 1334-1342.

Jusup, M., T. Sousa, T. Domingos, V. Labinac, N. Marn, Z. Wang and T. Klanjscek (2017): Physics of metabolic organization. Physics of Life Reviews 20: 1-39.

Kattwinkel, M. and M. Liess (2014): Competition matters: Species interactions prolong the long-term effects of pulsed toxicant stress on populations. Environmental Toxicology and Chemistry 33(7): 1458-1465.

Kattwinkel, M., P. Reichert, J. Ruegg, M. Liess and N. Schuwirth (2016): Modeling Macroinvertebrate Community Dynamics in Stream Mesocosms Contaminated with a Pesticide. Environmental Science & Technology 50(6): 3165-3173.

Kishi, M. J., S. Ito, B. A. Megrey, K. A. Rose and F. E. Werner (2011): A review of the NEMURO and NEMURO.FISH models and their application to marine ecosystem investigations. Journal of Oceanography 67(1): 3-16.

Kishi, M. J., M. Kashiwai, D. M. Ware, B. A. Megrey, D. L. Eslinger, F. E. Werner, M. Noguchi-Aita, T. Azumaya, M. Fujii, S. Hashimoto, D. J. Huang, H. Iizumi, Y. Ishida, S. Kang, G. A. Kantakov, H. C. Kim, K. Komatsu, V. V. Navrotsky, S. L. Smith, K. Tadokoro, A. Tsuda, O. Yamamura, Y. Yamanaka, K. Yokouchi, N. Yoshie, J. Zhang, Y. I. Zuenko and V. I. Zvalinsky (2007): NEMURO - a lower trophic level model for the North Pacific marine ecosystem. Ecological Modelling 202(1-2): 12-25.

Klanjscek, T., H. Caswell, M. G. Neubert and R. M. Nisbet (2006): Integrating dynamic energy budgets into matrix population models. Ecological Modelling 196(3-4): 407-420.

Klanjscek, T., R. M. Nisbet, H. Caswell and M. G. Neubert (2007): A model for energetics and bioaccumulation in marine mammals with applications to the right whale. Ecological Applications 17(8): 2233-2250.

Klanjscek, T., R. M. Nisbet, J. H. Priester and P. A. Holden (2012): Modeling physiological processes that relate toxicant exposure and bacterial population dynamics. Plos One 7(2).

Klanjscek, T., R. M. Nisbet, J. H. Priester and P. A. Holden (2013): Dynamic energy budget approach to modeling mechanisms of CdSe quantum dot toxicity. Ecotoxicology 22(2): 319-330.

Klok, C. (2008): Gaining insight in the interaction of zinc and population density with a combined dynamic energy budget and population model. Environmental Science & Technology 42: 8803-8808.

Klok, C. and A. M. De Roos (1996): Population level consequences of toxicological influences on individual growth and reproduction in *Lumbricus rubellus* (Lumbricidae, Oligochaeta). Ecotoxicology and Environmental Safety 33: 118-127.

Klok, C., A. M. De Roos, J. C. Y. Marinissen, H. M. Baveco and W. C. Ma (1997): Assessing the effects of abiotic environmental stress on population growth in *Lumbricus rubellus* (Lubricidae, Oligochaeta). Soil Biology & Biochemistry 29(3-4): 287-293.

Klok, C., M. Holmstrup and C. Damgaard (2007): Extending a combined dynamic energy budget matrix population model with a bayesian approach to assess variation in the intrinsic rate of population increase. An example in the earthworm *Dendrobaena octaedra*. Environmental Toxicology and Chemistry 26(11): 2383-2388.

Klok, C., A. Van der Hout and J. Bodt (2006): Population growth and development of the earthworm Lumbricus rubellus in a polluted field soil: Possible consequences for the godwit (Limosa limosa). Environmental Toxicology and Chemistry 25(1): 213-219.

Knäbe, S. (2013): Determination of Residues of Folpet and its Metabolite Phthalimide in Ground Vegetation after four Applications of Folpan 80 WDG in Vineyards in Germany 2012. Eurofins Agroscience Services EcoChem GmbH; Eutinger Str. 24, D-75223 Niefern-Öschelbronn, Germany. Report No. S12-01835. Sponsor: ADAMA Makhteshim Ltd., P.O.Box 60, Industrial Zone Beer-Sheva, 84100, Israel. Sponsor Ref. No. R-30368.

Knillmann, S., P. Orlinskiy, O. Kaske, K. Foit and M. Liess (2018): Indication of pesticide effects and recolonization in streams. Science of The Total Environment 630: 1619-1627.

Knillmann, S., N. C. Stampfli, M. A. Beketov and M. Liess (2012a): Intraspecific competition increases toxicant effects in outdoor pond microcosms. Ecotoxicology 21(7): 1857-1866.

Knillmann, S., N. C. Stampfli, Y. A. Noskov, M. A. Beketov and M. Liess (2012b): Interspecific competition delays recovery of Daphnia spp. populations from pesticide stress. Ecotoxicology 21(4): 1039-1049.

Koelmans, A. A., A. Van der Heijde, L. M. Knijff and R. H. Aalderink (2001): Integrated modelling of eutrophication and organic contaminant fate & effects in aquatic ecosystems. A review. Water Research 35(15): 3517-3536.

Koenraadt, C. J. M., S. Majambere, L. Hemerik and W. Takken (2004): The effects of food and space on the occurrence of cannibalism and predation among larvae of Anopheles gambiae s.l. Entomologia Experimentalis et Applicata 112(2): 125-134.

Koh, H. L., T. G. Hallam and H. L. Lee (1997): Combined effects of environmental and chemical stressors on a model Daphnia population. Ecological Modelling 103(1): 19-32.

Könemann, H. and K. Van Leeuwen (1980): Toxicokinetics in fish: accumulation and elimination of six chlorobenzenes by guppies. Chemosphere 9(1): 3-19.

Kooi, B. W. (2003): Numerical bifurcation analysis of ecosystems in a spatially homogeneous environment. Acta Biotheoretica 51(3): 189-222.

Kooi, B. W., D. Bontje and M. Liebig (2008a): Model analysis of a simple aquatic ecosystems with sublethal toxic effects. Mathematical Biosciences and Engineering 5(4): 771-787.

Kooi, B. W., D. Bontje, G. A. K. van Voorn and S. A. L. M. Kooijman (2008b): Sublethal toxic effects in a simple aquatic food chain. Ecological Modelling 212(3-4): 304-318.

Kooijman, S. A. L. M. (2000). Dynamic Energy and Mass Budget in Biological Systems. Cambridge University Press; Cambirdge, UK.

Kooijman, S. A. L. M. (2001): Quantitative aspects of metabolic organization: a discussion of concepts. Philosophical Transactions of the Royal Society of London B 356: 331-349.

Kooijman, S. A. L. M. (2010). Dynamic energy budget theory for metabolic organisation. Cambridge University Press; Cambridge, UK.

Kooijman, S. A. L. M. and J. J. M. Bedaux (1996a). The analysis of aquatic toxicity data. VU University Press; Amsterdam, The Netherlands.

Kooijman, S. A. L. M. and J. J. M. Bedaux (1996b): Analysis of toxicity tests on *Daphnia* survival and reproduction. Water Research 30(7): 1711-1723.

Kooijman, S. A. L. M. and J. J. M. Bedaux (1996c): Analysis of toxicity tests on fish growth. Water Research 30(7): 1633-1644.

Kooijman, S. A. L. M., A. O. Hanstveit and N. Nyholm (1996): No-effect concentrations in algal growth inhibition tests. Water Research 30(7): 1625-1632.

Kooijman, S. A. L. M. and J. A. J. Metz (1984a): On the dynamics of chemically stressed populations: the deduction of population consequences from effects on individuals. Ecotoxicology and Environmental Safety 8: 254-274.

Kooijman, S. A. L. M. and J. A. J. Metz (1984b): On the dynamics of chemically stressed populations: the deduction of population consequences from effects on individuals. Ecotoxicology and Environmental Safety 8(3): 254-274.

Krishnan, K. and T. Peyret (2009). Physiologically based toxicokinetic (PBTK) modeling in ecotoxicology. Book chapter in: Ecotoxicology modeling. J. Devillers. Springer; New York, USA: 145-175.

Kusk, K. O. (1996): Bioavailability and effect of pirimicarb on Daphnia magna in a laboratory freshwater/sediment system. Archives of Environmental Contamination and Toxicology 31(2): 252-255.

Legind, C. N., C. M. Kennedy, A. Rein, N. Snyder and S. Trapp (2011): Dynamic plant uptake model applied for drip irrigation of an insecticide to pepper fruit plants. Pest Management Science 67(5): 521-527.

Lei, B., S. Huang, M. Qiao, T. Li and Z. Wang (2008): Prediction of the environmental fate and aquatic ecological impact of nitrobenzene in the Songhua River using the modified AQUATOX model. Journal of Environmental Sciences 20(7): 769-777.

Lien, G. J., J. W. Nichols, J. M. McKim and C. A. Gallinat (1994): Modeling the accumulation of three waterborne chlorinated ethanes in fathead minnows (*Pimephales promelas*): a physiologically based approach. Environmental Toxicology and Chemistry 13(7): 1195-1205.

Liess, M. and K. Foit (2010a): Intraspecific competition delays recovery of population structure. Aquatic Toxicology 97(1): 15-22.

Liess, M. and K. Foit (2010b): Intraspecific competition delays recovery of population structure. Aquat Toxicol 97(1): 15-22.

Liess, M., K. Foit, A. Becker, E. Hassold, I. Dolciotti, M. Kattwinkel and S. Duquesne (2013): Culmination of low-dose pesticide effects. Environmental Science & Technology 47(15): 8862-8868.

Liess, M., K. Foit, S. Knillmann, R. B. Schaefer and H.-D. Liess (2016a): Predicting the synergy of multiple stress effects. Scientific Reports 6.

Liess, M., K. Foit, S. Knillmann, R. B. Schafer and H. D. Liess (2016b): Predicting the synergy of multiple stress effects. Scientific Reports 6: 8.

Liess, M., N. V. Gerner and B. J. Kefford (2017): Metal toxicity affects predatory stream invertebrates less than other functional feeding groups. Environmental Pollution 227: 505-512.

Liess, M., S. Henz and N. Shahid (2020): Modeling the synergistic effects of toxicant mixtures. Environmental Sciences Europe 32(1): 119.

Liess, M., L. Liebmann, P. Vormeier, O. Weisner, R. Altenburger, D. Borchardt, W. Brack, A. Chatzinotas, B. Escher, K. Foit, R. Gunold, S. Henz, K. L. Hitzfeld, M. Schmitt-Jansen, N. Kamjunke, O. Kaske, S. Knillmann, M. Krauss, E. Küster, M. Link, M. Lück, M. Möder, A. Müller, A. Paschke, R. B. Schäfer, A. Schneeweiss, V. C. Schreiner, T. Schulze, G. Schüürmann, W. von Tümpling, M. Weitere, J. Wogram and T. Reemtsma (2021): Pesticides are the dominant stressors for vulnerable insects in lowland streams. Water Research 201: 117262.

Liess, M., R. B. Schafer and C. A. Schriever (2008): The footprint of pesticide stress in communities - species traits reveal community effects of toxicants. Sci Total Environ 406(3): 484-490.

Liess, M. and R. Schulz (1996): Chronic effects of short-term contamination with the pyrethroid insecticide fenvalerate on the caddisfly Limnephilus lunatus. Hydrobiologia 324(2): 99-106.

Liess, M. and R. Schulz (1999): Linking insecticide contamination and population response in an agricultural stream. Environmental Toxicology and Chemistry 18(9): 1948–1955.

Liess, M., R. Schulz, M. H. D. Liess, B. Rother and R. Kreuzig (1999): Determination of insecticide contamination in agricultural headwater streams. Water Research 33(1): 239-247.

Liess, M. and P. C. von der Ohe (2005): Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry 24(4): 954–965.

Lika, K., M. R. Kearney, V. Freitas, H. W. Van der Veer, J. Van der Meer, J. W. M. Wijsman, L. Pecquerie and S. A. L. M. Kooijman (2011): The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model I: Philosophy and approach. Journal of Sea Research 66(4): 270-277.

Lika, K. and R. M. Nisbet (2000): A Dynamic Energy Budget model based on partitioning of net production. Journal of Mathematical Biology 41(4): 361-386.

Liu, C., A. J. Bednarska, R. M. Sibly, R. C. Murfitt, P. Edwards and P. Thorbek (2014): Incorporating toxicokinetics into an individual-based model for more realistic pesticide exposure estimates: A case study of the wood mouse. Ecological Modelling 280: 30-39.

Liu, C., R. M. Sibly, V. Grimm and P. Thorbek (2013): Linking pesticide exposure and spatial dynamics: An individual-based model of wood mouse (Apodemus sylvaticus) populations in agricultural landscapes. Ecological Modelling 248: 92-102.

Llandres, A. L., G. M. Marques, J. L. Maino, S. A. L. M. Kooijman, M. R. Kearney and J. Casas (2015): A dynamic energy budget for the whole life-cycle of holometabolous insects. Ecological Monographs 85(3): 353-371.

Lombardo, A., A. Franco, A. Pivato and A. Barausse (2015): Food web modeling of a river ecosystem for risk assessment of down-the-drain chemicals: A case study with AQUATOX. Science of the Total Environment 508: 214-227.

Loos, M., A. M. J. Ragas, A. J. Hendriks and R. Plasmeijer (2008): Ecological and Spatially explicit Cumulative & Exposure model (Eco-SpaCE), an individual-based ecological exposure model for terrestrial vertebrates in an object-oriented programming platform. NERI.

Loos, M., A. M. J. Ragas, R. Plasmeijer, A. M. Schipper and A. J. Hendriks (2010): Eco-SpaCE: An object-oriented, spatially explicit model to assess the risk of multiple environmental stressors on terrestrial vertebrate populations. Science of the Total Environment 408(18): 3908-3917.

Lopes, C., A. R. R. Pery, A. Chaumot and S. Charles (2005): Ecotoxicology and population dynamics: using DEBtox models in a Leslie modeling approach. Ecological Modelling 188: 30-40.

Loucks, D. P. and E. van Beek (2017). Water Resource Systems Planning and Management - An Introduction to Methods, Models, and Applications. Springer Nature, Open Deltares and UNESCO-IHE; Springer International Publishing AG, Gewerbestrasse 11, 6330 Cham, Switzerland.

Macdonald, D. W. and P. Barrett (2005). Collins Field Guide – Mammals of Britain and Europe. (Reissue) ed. Collins; London, UK.

Mackay, D., H. Puig and L. S. McCarty (1992): An equation describing the time course and variability in uptake and toxicity of narcotic chemicals to fish. Environmental Toxicology and Chemistry 11: 941-951.

Mancini, J. L. (1983): A method for calculating effects, on aquatic organisms, of time varying concentrations. Water Research 17(10): 1355-1362.

Marques, G. M., S. Augustine, K. Lika, L. Pecquerie, T. Domingos and S. A. L. M. Kooijman (2018): The AmP project: comparing species on the basis of dynamic energy budget parameters. Plos Computational Biology 14(5): e1006100.

Martin, B., T. Jager, R. M. Nisbet, T. G. Preuss and V. Grimm (2014): Limitations of extrapolating toxic effects on reproduction to the population level. Ecological Applications 24(8): 1972-1983.

Martin, B. T., T. Jager, R. M. Nisbet, T. G. Preuss and V. Grimm (2013a): Predicting population dynamics from the properties of individuals: a cross-level test of Dynamic Energy Budget theory. American Naturalist 181(4): 506-519.

Martin, B. T., T. Jager, R. M. Nisbet, T. G. Preuss, M. Hammers-Wirtz and V. Grimm (2013b): Extrapolating ecotoxicological effects from individuals to populations: a generic approach based on Dynamic Energy Budget theory and individual-based modeling. Ecotoxicology 22: 574-583.

Martin, B. T., E. I. Zimmer, V. Grimm and T. Jager (2012): Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. Methods in Ecology and Evolution 3(2): 445-449.

Massarin, S., R. Beaudouin, F. Zeman, M. Floriani, R. Gilbin, F. Alonzo and A. R. R. Pery (2011): Biology-based modeling to analyze uranium toxicity data on *Daphnia magna* in a multigeneration study. Environmental Science & Technology 45(9): 4151-4158.

Mastitsky, S. E., M. Ebeling, N. Kreling, P. Thorbek and W. Schmitt (2013): Investigating population relevance of the effects observed in standard rat reproduction studies. Proceedings of the SETAC Europe 23rd Annual Meeting. Glasgow, UK.

McKnight, U. S., S. G. Funder, J. J. Rasmussen, M. Finkel, P. J. Binning and P. L. Bjerg (2010): An integrated model for assessing the risk of TCE groundwater contamination to human receptors and surface water ecosystems. Ecological Engineering 36(9): 1126-1137.

McMahon, D. P., M. E. Natsopoulou, V. Doublet, M. Fürst, S. Weging, M. J. F. Brown, A. Gogol-Döring and R. J. Paxton (2016): Elevated virulence of an emerging viral genotype as a driver of honeybee loss. Proceedings of the Royal Society B 283(1833).

Meli, M., A. Auclerc, A. Palmqvist, V. E. Forbes and V. Grimm (2013): Population-level consequences of spatially heterogeneous exposure to heavy metals in soil: An individual-based model of springtails. Ecological Modelling 250: 338-351.

Meli, M., A. Palmqvist and V. E. Forbes (2014a): Implications of interacting microscale habitat heterogeneity and disturbance events on *Folsomia candida* (Collembola) population dynamics: A modeling approach. Environmental Toxicology and Chemistry 33(7): 1508-1516.

Meli, M., A. Palmqvist, V. E. Forbes, J. Groeneveld and V. Grimm (2014b): Two pairs of eyes are better than one: Combining individual-based and matrix models for ecological risk assessment of chemicals. Ecological Modelling 280: 40-52.

Miller, F. J., P. M. Schlosser and D. B. Janszen (2000): Haber's rule: a special case in a family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint. Toxicology 149(1): 21-34.

Miller, R. J., H. S. Lenihan, E. B. Muller, N. Tseng, S. K. Hanna and A. A. Keller (2010): Impacts of metal oxide nanoparticles on marine phytoplankton. Environmental Science & Technology 44(19): 7329-7334.

Moore, M. V. (1988): Density-dependent predation of early instar Chaoborus feeding on multispecies prey assemblages. Limnology and Oceanography(33): 256-268.

Muir, D., A. Fisk and M. Kwan (2001): Temporal trends of persistent organic pollutants and metals in ringed seals from the Canadian Arctic. Synopsis of Research Conducted Under the 2000-2001 Northern Contaminants Program: 208-214.

Muller, E. B., S. A. L. M. Kooijman, P. J. Edmunds, F. J. Doyle and R. M. Nisbet (2009): Dynamic energy budgets in syntrophic symbiotic relationships between heterotrophic hosts and photoautotrophic symbionts. Journal of Theoretical Biology 259(1): 44-57.

Muller, E. B. and R. M. Nisbet (2014): Dynamic energy budget modeling reveals the potential of future growth and calcification for the coccolithophore *Emiliania huxleyi* in an acidified ocean. Global Change Biology 20(6): 2031-2038.

Muller, E. B., R. M. Nisbet and H. A. Berkley (2010): Sublethal toxicant effects with dynamic energy budget theory: model formulation. Ecotoxicology 19(1): 48-60.

Münze, R., C. Hannemann, P. Orlinskiy, R. Gunold, A. Paschke, K. Foit, J. M. Becker, O. Kaske, E. Paulsson, M. Peterson, H. Jernstedt, J. Kreuger, G. Schüürmann and M. Liess (2017): Pesticides from wastewater treatment plant effluents affect invertebrate communities. Science of the Total Environment 599-600: 387-399.

Münze, R., P. Orlinskiy, R. Gunold, A. Paschke, O. Kaske, M. A. Beketov, M. Hundt, C. Bauer, G. Schüürmann, M. Möder and M. Liess (2015): Pesticide impact on aquatic invertebrates identified with Chemcatcher® passive samplers and the SPEARpesticides index. Science of the Total Environment 537: 69-80.

Naito, W., K. Miyamoto, J. Nakanishi, S. Masunaga and S. M. Bartell (2003): Evaluation of an ecosystem model in ecological risk assessment of chemicals. Chemosphere 53(4): 363-375.

Nichols, J. W., J. M. McKim, M. E. Andersen, M. L. Gargas, H. J. Clewell and R. J. Erickson (1990): A physiologicall based toxicokinetic model for the uptake and disposition of waterborne organic chemicals in fish. Toxicology and Applied Pharmacology 106: 433-447.

Niethammer, F. and F. Krapp (1982). Microtus agrestis (Linnaeus, 1761) – Erdmaus. Book chapter in: Handbuch der Säugetiere Europas Vol. 2/I, Nagetiere II. F. Niethammer and F. Krapp. Akademische Verlagsgesellschaft Wiesbaden; Wiesbaden, Germany: 349-373.

Niewersch, C. (2005). Entwicklung eines ökotoxikologischen Tests mit der Büschelmücke Chaoborus crystallinus (Diptera: Chaoboridae). Diploma Thesis Thesis, RWTH Aachen University.

Nisbet, R. M., E. McCauley and L. R. Johnson (2010): Dynamic energy budget theory and population ecology: lessons from *Daphnia*. Philosophical Transactions of the Royal Society B-Biological Sciences 365(1557): 3541-3552.

Nisbet, R. M., E. B. Muller, K. Lika and S. A. L. M. Kooijman (2000): From molecules to ecosystems through dynamic energy budget models. Journal of Animal Ecology 69: 913-926.

Noël, H. L., S. P. Hopkin, T. H. Hutchinson, T. D. Williams and R. M. Sibly (2006): Population growth rate and carrying capacity for springtails *Folsomia candida* exposed to ivermectin. Ecological Applications 16(2): 656-665.

Noonburg, E. G., R. M. Nisbet and T. Klanjscek (2010): Effects of life history variation on vertical transfer of toxicants in marine mammals. Journal of Theoretical Biology 264(2): 479-489.

Nyman, A. M., K. Schirmer and R. Ashauer (2012): Toxicokinetic-toxicodynamic modelling of survival of *Gammarus pulex* in multiple pulse exposures to propiconazole: model assumptions, calibration data requirements and predictive power. Ecotoxicology 21(7): 1828-1840.

OECD (2006). Current approaches in the statistical analysis of ecotoxicity data: a guidance to application, Series on Testing and Assessment, No. 54. Organisation for Economic Cooperation and Development (OECD); Paris, France.

OECD (2012). OECD guideline for testing of chemicals 305. Bioaccumulation in fish: aqueous and dietary exposure. Organisation for Economic Cooperation and Development (OECD); Paris, France.

Oksanen, T. A., M. Koivula, E. Koskela and T. Mappes (2007): The cost of reproduction induced by body size at birth and breeding density. Evolution 61(12): 2822-2831.

Orlinskiy, P., R. Münze, M. A. Beketov, R. Gunold, A. Paschke, S. Knillmann and M. Liess (2015): Forested headwaters mitigate pesticide effects on macroinvertebrate communities in streams: Mechanisms and quantification. Science of the Total Environment 524-525: 115-123.

Ouin, A., G. Paillat, A. Butet and F. Burel (2000): Spatial dynamics of wood mouse (Apodemus sylvaticus) in an agricultural landscape under intensive use in the Mont Saint Michel Bay (France). Agriculture, Ecosystems & Environment 78(2): 159-165.

Pablos, M. V., S. Boleas and J. V. Tarazona (1998): Use of MFU-galactoside enzymatic activity as ecotoxicological endpoint on rainbow trout red blood cells. Bulletin of Environmental Contamination and Toxicology 61(6): 786-792.

Park, R. A. and J. S. Clough (2014): AQUATOX (Release 3.1 plus) - Modelling environmental fate and ecologcial effects in aquatic ecosystems. Volume 2: Technical Documentation. United States Environmental Protection Agency (USEPA). Office of Water (4305). Office of Science and Technology; Washington DC 20460 , USA. Report No. EPA-820-R-14-007.

Park, R. A. and J. S. Clough (2018): AQUATOX (Release 3.2) - Modelling environmental fate and ecologcial effects in aquatic ecosystems. Volume 2: Technical Documentation. United States Environmental Protection Agency (USEPA). Office of Research and Development. Office of Science and Technology; Washington DC 20460, USA. Report No. EPA/600/B-18/241.

Park, R. A., J. S. Clough and M. C. Wellman (2008): AQUATOX: Modeling environmental fate and ecological effects in aquatic ecosystems. Ecological Modelling 213(1): 1-15.

Paterson, S., D. Mackay and C. McFarlane (1994): A model of organic chemical uptake by plants from soil and the atmosphere. Environmental Science & Technology 28(13): 2259-2266.

Pedersen, S., H. Selck, D. Salvito and V. Forbes (2009): Effects of the polycyclic musk HHCB on individual- and population-level endpoints in *Potamopyrgus antipodarum*. Ecotoxicology and Environmental Safety 72(4): 1190-1199.

Peeters, F., J. Li, D. Straile, K. O. Rothhaupt and J. Vijverberg (2010): Influence of low and decreasing food levels on *Daphnia*-algal interactions: Numerical experiments with a new dynamic energy budget model. Ecological Modelling 221(22): 2642-2655.

Persson, L. and A. M. de Roos (2006): Size-Structured Interactions and the Dynamics of Aquatic Systems. Polish Journal of Ecology 54(4).

Pickering, F. (2016): Beta-cyfluthrin (Bulldock 25EC). Extended laboratory test to assess the sensitivity of phantom midge larvae (Chaoborus sp.) over 96 hours. CEA (ADAS UK Ltd); Report No. XCE2012_005.

Pieters, B. J., T. Jager, M. H. S. Kraak and W. Admiraal (2006): Modeling responses of *Daphnia magna* to pesticide pulse exposure under varying food conditions: intrinsic versus apparent sensitivity. Ecotoxicology 15: 601-608.

Pieters, B. J. and M. Liess (2006): Maternal nutritional state determines the sensitivity of Daphnia magna offspring to short-term Fenvalerate exposure. Aquatic Toxicology 76(3): 268-277.

Pisa, L. W., V. Amaral-Rogers, L. P. Belzunces, J. M. Bonmatin, C. A. Downs, D. Goulson, D. P. Kreutzweiser, C. Krupke, M. Liess, M. McField, C. A. Morrissey, D. A. Noome, J. Settele, N. Simon-Delso, J. D. Stark, J. P. Van der Sluijs, H. Van Dyck and M. Wiemers (2015): Effects of neonicotinoids and fipronil on non-target invertebrates. Environmental Science and Pollution Research 22(1): 68-102.

Preuss, T. G. (2015): Virtual pulse exposure reproduction study of Bromoxynil-octanoate on Daphnia magna with the IDamP model. Report No. EnSa-14-1051. Document No. M-497765-02-1.

Preuss, T. G., M. Hammers-Wirtz, U. Hommen, M. N. Rubach and H. T. Ratte (2009a): Development and validation of an individual based Daphnia magna population model: The influence of crowding on population dynamics. Ecological Modelling 220: 310-329.

Preuss, T. G., M. Hammers-Wirtz and H. T. Ratte (2010a): The potential of individual based population models to extrapolate effects measured at standardized test conditions to relevant environmental conditions-an example for 3,4-dichloroaniline on Daphnia magna. Journal of Environmental Monitoring 12(11): 2070-2079.

Preuss, T. G., M. Hammers-Wirtz and H. T. Ratte (2010b): The potential of individual based population models to extrapolate effects measured at standardized test conditions to relevant environmental conditions - an example for 3,4-dichloroaniline on *Daphnia magna*. Journal of Environmental Monitoring 12(11): 2070-2079.

Preuss, T. G., T. Strauss, M. Hammers-Wirtz and H. T. Ratte (2009b): Modelling effects of pirimicarb on Daphnia magna populations under simulated field conditions. Research Institute for Ecosystem Analysis and Assessment - gaiac; Institute for Environmental Research (Biology V), RWTH Aachen, Worringerweg 1, D- 52056 Aachen, Germany. Report No. T009915-06. Sponsor: Syngenta, Jealott's Hill, Bracknell, United Kingdom. Sponsor Ref. No. PP62_10059.

Preziosi, D. V. and R. A. Pastorok (2008): Ecological food web analysis for chemical risk assessment. Sci Total Environ 406(3): 491-502.

Railsback, S. F. and V. Grimm (2011). Agent-Based and Individual-Based Modeling - A Practical Introduction. Princeton University Press.

Raimondo, S., D. N. Vivian and M. G. Barron (2010): Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual. Version 1.1. USEPA; Gulf Breeze, FL. Report No. EPA/600/R-10/004.

Ramcharan, C. W., D. J. McQueen, A. Pérez-Fuentetaja, N. D. Yan, E. Demers and J. Rusak (2001): Analyses of lake food webs using individual-based models to estimate Chaoborus production and consumption. Arch. Hydrobiol. Spec. Issues Adv. Limnol. 56: 101-126.

Ramskov, T. and V. E. Forbes (2008): Life history and population dynamics of the opportunistic polychaete *Capitella* sp I in relation to sediment organic matter. Marine Ecology Progress Series 369: 181-192.

Rashleigh, B., M. C. Barber and D. M. Walters (2009): Foodweb modeling for polychlorinated biphenyls (PCBs) in the Twelvemile Creek Arm of Lake Hartwell, South Carolina, USA. Ecological Modelling 220(2): 254-264.

Reed, M., T. Alvarez, S. Chelinho, V. Forbes, A. Johnston, M. Meli, F. Voss and R. Pastorok (2016): A Risk Assessment Example for Soil Invertebrates Using Spatially Explicit Agent-Based Models. Integrated Environmental Assessment and Management 12(1): 58-66.

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal of the European Union, L 309, p. 1. 24.11.2009.

Reichenberger, S., M. Bach, A. Skitschak and H. G. Frede (2007): Mitigation strategies to reduce pesticide inputs into ground- and surface water and their effectiveness; a review. Sci Total Environ 384(1-3): 1-35.

Reichstein, H. (1960): Untersuchungen zum Aktionsraum und zum Revierverhalten der Feldmaus. Z. Säugetierkd 25: 150-168.

Rein, A., C. N. Legind and S. Trapp (2011): New concepts for dynamic plant uptake models. Sar and Qsar in Environmental Research 22(1-2): 191-215.

Richter, O., B. Diekkrüger and P. Nörtersheuser (1996). Environmental Fate Modelling of Pesticides: From the Laboratory to the Field Scale. VCH Verlagsgesellschaft mbH; 69451 Weinheim, Germany.

RIFCON (2017): TRACE document for SpringSimPy – RIFCON population model for the springtail, *Folsomia candida* (Version 1.0). RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany.

RIFCON (2018): TRACE document for eVole - RIFCON Common Vole Population Model (Version 3.0). RIFCON GmbH; Goldbeckstr. 13, D-69493 Hirschberg, Germany.

Rinke, K. and J. Vijverberg (2005): A model approach to evaluate the effect of temperature and food concentration on individual life-history and population dynamics of Daphnia. Ecological Modelling 186(3): 326-344.

Roessink, I., J. D. Belgers, S. J. Crum, P. J. van den Brink and T. C. Brock (2006): Impact of triphenyltin acetate in microcosms simulating floodplain lakes. II. Comparison of species sensitivity distributions between laboratory and semi-field. Ecotoxicology 15(5): 411-424.

Rubach, M. N., R. Ashauer, S. J. Maund, D. J. Baird and P. J. Van den Brink (2010): Toxicokinetic variation in 15 freshwater arthropod species exposed to the insecticide chlorpyrifos. Environmental Toxicology and Chemistry 29(10): 2225-2234.

Rubin, Y. (1985): Folpan, Teratology study in the rabbit. Life Science Research Israel Ltd.; PO Box 139, Ness Ziona, Israel. Report No. MAK/051/FOL, R-3684.

Rubin, Y. (1986): Folpan, Two-generation reproduction study in the rat. Life Science Research Israel Ltd.; PO Box 139, Ness Ziona, Israel. Report No. MAK/052/FOL, R-4347.

Rumkee, J. C. O., M. A. Becher, P. Thorbek, P. J. Kennedy and J. L. Osborne (2015): Predicting Honeybee Colony Failure: Using the BEEHAVE Model to Simulate Colony Responses to Pesticides. Environmental Science & Technology 49(21): 12879-12887.

Russo, R., J. M. Becker and M. Liess (2018): Sequential exposure to low levels of pesticides and temperature stress increase toxicological sensitivity of crustaceans. Science of The Total Environment 610–611: 563-569.

Sakamoto, N., S. Saito, T. Hirose, M. Suzuki, S. Matsuo, K. Izumi, T. Nagatomi, H. Ikegami, K. Umeda, K. Tsushima and N. Matsuo (2004): The discovery of pyridalyl: a novel insecticidal agent for controlling lepidopterous pests. Pest Management Science 60(1): 25-34.

Salice, C. J. and T. J. Miller (2003): Population-level responses to long-term cadmium exposure in two strains of the freshwater gastropod *Biomphalaria glabrata*: Results from a life-table response experiment. Environmental Toxicology and Chemistry 22(3): 678-688.

Sánchez-Bayo, F., K. Goka and D. Hayasaka (2016): Contamination of the Aquatic Environment with Neonicotinoids and its Implication for Ecosystems. Frontiers in Environmental Science 4(71).

SANCO/10329/2002 rev. 2 (2002). Draft Working Document - Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC. European Commission Health & Consumer Protection Directorate-General. Directorate E - Food Safety: plant health, animal health and welfare, international questions. E1 - Plant health.

Schaefer, R. B., T. Caquet, K. Siimes, R. Mueller, L. Lagadic and M. Liess (2007): Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe. Sci Total Environ 382(2-3): 272-285.

Schaefer, R. B., B. J. Kefford, L. Metzeling, M. Liess, S. Burgert, R. Marchant, V. Pettigrove, P. Goonan and D. Nugegoda (2011): A trait database of stream invertebrates for the ecological risk assessment of single and combined effects of salinity and pesticides in South-East Australia. Science of the Total Environment 409(11): 2055-2063.

Schaefer, R. B., P. C. von der Ohe, J. Rasmussen, B. J. Kefford, M. A. Beketov, R. Schulz and M. Liess (2012): Thresholds for the Effects of Pesticides on Invertebrate Communities and Leaf Breakdown in Stream Ecosystems. Environmental Science & Technology 46(9): 5134-5142.

Schanné, C. and J. van der Kolk (2001): [14C]-Deltamethrin Formulated as Emulsifiable Concentrate (25 g/L Deltamethrin): Outdoor Aquatic Microcosm Study of the Ecological Effects and Environmental Fate. SpringBorn Laboratories; Horn, Switzerland.

Schipper, A. M., M. Loos, A. M. J. Ragas, J. P. C. Lopes, B. T. Nolte, S. Wijnhoven and R. Leuven (2008): Modeling the influence of environmental heterogeneity on heavy metal exposure concentrations for terrestrial vertebrates in river floodplains. Environmental Toxicology and Chemistry 27(4): 919-932.

Schletterer, M., L. Fureder, V. V. Kuzovlev and M. A. Beketov (2010): Testing the coherence of several macroinvertebrate indices and environmental factors in a large lowland river system (Volga River, Russia). Ecological Indicators 10(6): 1083-1092.

Schmitt, W., D. Auteri, F. Bastiansen, M. Ebeling, C. Liu, R. Luttik, S. Mastitsky, D. Nacci, C. Topping and M. Wang (2015): An Example of Population-Level Risk Assessments for Small Mammals Using Individual-Based Population Models. Integrated Environmental Assessment and Management 12(1): 46-57.

Schmitt, W., E. Bruns, M. Dollinger and P. Sowig (2013): Mechanistic TK/TD-model simulating the effect of growth inhibitors on *Lemna* populations. Ecological Modelling 255: 1-10.

Schmolke, A., P. Thorbek, D. L. DeAngelis and V. Grimm (2010): Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology & Evolution 25(8): 479-486.

Schuwirth, N. and P. Reichert (2013): Bridging the gap between theoretical ecology and real ecosystems: modeling invertebrate community composition in streams. Ecology 94(2): 368-379.

Sequeira, A. and J. G. Ferreira (2005): ECASA - Model description template for ECOWIN 2000.

Sevim, S. (2012). Beurteilung des Wiedererholungspotenzials emergierender Insekten am Beispiel der Büschelmücke Chaoborus crystallinus in Mesokosmenstudien unter Berücksichtigung der Populationsstruktur (Metapopulation) und der Verbreitungsmechanismen. PhD Thesis Thesis, RWTH Aachen University.

Sibly, R. M., V. Grimm, B. T. Martin, A. S. A. Johnston, K. Kulakowska, C. J. Topping, P. Calow, J. Nabe-Nielsen, P. Thorbek and D. L. DeAngelis (2013): Representing the acquisition and use of energy by individuals in agent-based models of animal populations. Methods in Ecology and Evolution 4(2): 151-161.

Sinko, J. W. and W. Streifer (1969): Applying models incorporating age-size structure of a population to *Daphnia*. Ecology 50(4): 608-&.

Smit, M. G. D., B. J. Kater, R. G. Jak and M. J. van den Heuvel-Greve (2006): Translating bioassay results to field population responses using a Leslie-matrix model for the marine amphipod *Corophium volutator*. Ecological Modelling 196(3-4): 515-526.

Smyth, M. (1966): Winter Breeding in Woodland Mice, Apodemus Sylvaticus, and Voles, Clethrionomys Glareolus and Microtus Agrestis, Near Oxford. Journal of Animal Ecology 35(3): 471-485.

Soetaert, K. and P. M. J. Herman (2009). A Practical Guide to Ecological Modelling Using R as a Simulation Platform. Springer Science + Business Media B.V.; Netherlands Institute of Ecology, Yerseke, The Netherlands.

Sourisseau, S., A. Bassères, F. Périé and T. Caquet (2008): Calibration, validation and sensitivity analysis of an ecosystem model applied to artificial streams. Water Research 42(4): 1167-1181.

Sousa, T., T. Domingos and S. A. L. M. Kooijman (2008): From empirical patterns to theory: a formal metabolic theory of life. Philosophical Transactions of the Royal Society B-Biological Sciences 363(1502): 2453-2464.

Sousa, T., T. Domingos, J. C. Poggiale and S. A. L. M. Kooijman (2010): Dynamic energy budget theory restores coherence in biology. Philosophical Transactions of the Royal Society B-Biological Sciences 365(1557): 3413-3428.

Spacie, A. and J. L. Hamelink (1982): Alternative models for describing the bioconcentration of organics in fish. Environmental Toxicology and Chemistry 1: 309-320.

Spann, N., W. Goedkoop and W. Traunspurger (2015): Phenanthrene Bioaccumulation in the Nematode Caenorhabditis elegans. Environmental Science & Technology 49(3): 1842-1850.

Springborn (2006): [^{14}C]-dichlorophenyl S-1812 formulated as 10 EW (110g/kg): Effects and residue assessment for sediment-dwelling organisms exposed to natural water-sediment systems under realistic conditions, simulating 4 entries of the test item at 7 day intervals. Springborn Smither Laboratories (Europe) AG; Horn, Switzerland.

Spromberg, J. A. and J. P. Meador (2006): Relating chronic toxicity responses to population-level effects: A comparison of population-level parameters for three salmon species as a function of low-level toxicity. Ecological Modelling 199(3): 240-252.

Sprules, W. G. (1972): Effects of size-selective predation and food competition on high altitude zooplankton communities. Ecology 53(3): 375-386.

Stadnicka, J., K. Schirmer and R. Ashauer (2012): Predicting concentrations of organic chemicals in fish by using toxicokinetic models. Environmental Science & Technology 46(6): 3273-3280.

Steen Redeker, E., L. Bervoets and R. Blust (2004): Dynamic model for the accumulation of cadmium and zinc from water and sediment by the aquatic oligochaete, *Tubifex tubifex*. Environmental Science & Technology 38(23): 6193-6200.

Strauss, T. (2017): Description of the individual-based model "IBM Chaoborus population model" for the aquatic phantom midge Chaoborus crystallinus. Research Institute for Ecosystem Analysis and Assessment (gaiac).

Strauss, T., F. Gabsi, M. Hammer-Wirtz, P. Thorbek and T. G. Preuss (2017): The power of hybrid modelling: an example from aquatic ecosystems. Ecological Modelling 364: 77-88.

Strauss, T., D. Kulkarni, T. G. Preuss and M. Hammers-Wirtz (2016): The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

Strauss, T. and S. Norman (2017): Modelling Chaoborus crystallinus populations to simulate effects and recovery under beta-cyfluthrin exposure. gaiac - Research Institute for Ecosystem Analysis and Assessment; Kackertstr. 10, 52072 Aachen, Germany. Report No. R-37809. Sponsor: ADAMA Makhteshim Ldt., Beer-Sheva 84100, Israel.

Tattersall, F. H. and D. W. Macdonald (2003). Wood mouse in the arable ecosystem. Book chapter in: Conservation and Conflict: 82–96.

Thompson, R. S., T. D. Williams and J. F. Tapp (1989): Pirimicarb: Determination of chronic toxicity to Daphnia magna. Imperial Chemical Industries PLC; Brixham Laboratory, Freshwater Quarry, Brixham Devon TQ5 8BA, UK. Report No. BLB/B/3509. Sponsor: ICI Agrochemicals, Imperial Chemical Industries PLC, Fernhurst, Haslemere, Surrey GU27 3JE, UK. Sponsor Ref. No. R1106/B.

Thuiller, W. and T. Münkemüller (2010): Habitat suitability modeling. Effects of Climate Change on Birds: 77-85.

Tkadlec, E. and N. C. Stenseth (2001): A new geographical gradient in vole population dynamics. Proceedings: Biological Sciences 268(1476): 1547-1552.

Topping, C. J., P. S. Craig, F. de Jong, M. Klein, R. Laskowski, B. Manachini, S. Pieper, R. Smith, J. P. Sousa, F. Streissl, K. Swarowsky, A. Tiktak and T. van der Linden (2015): Towards a landscape scale management of pesticides: ERA using changes in modelled occupancy and abundance to assess long-term population impacts of pesticides. Science of the Total Environment 537: 159-169.

Topping, C. J., L. Dalby and F. Skov (2016): Landscape structure and management alter the outcome of a pesticide ERA: Evaluating impacts of endocrine disruption using the ALMaSS European Brown Hare model. Science of the Total Environment 541: 1477-1488.

Topping, C. J., T. Dalkvist and V. Grimm (2012): Post-Hoc Pattern-Oriented Testing and Tuning of an Existing Large Model: Lessons from the Field Vole. Plos One 7(9).

Topping, C. J., T. S. Hansen, T. S. Jensen, J. U. Jepsen, F. Nikolajsen and P. Odderskaer (2003): ALMaSS, an agent-based model for animals in temperate European landscapes. Ecological Modelling 167: 65-82.

Topping, C. J., T. T. Hoye, P. Odderskaer and N. J. Aebischer (2010a): A pattern-oriented modelling approach to simulating populations of grey partridge. Ecological Modelling 221: 729-737.

Topping, C. J., T. T. Hoye and C. R. Olesen (2010b): Opening the black box—Development, testing and documentation of a mechanistically rich agent-based model. Ecological Modelling 221(2): 245-255.

Topping, C. J., L. J. Kjaer, U. Hommen, T. T. Hoye, T. G. Preuss, R. M. Sibly and P. van Vliet (2014): RECOVERY BASED ON PLOT EXPERIMENTS IS A POOR PREDICTOR OF LANDSCAPE-LEVEL POPULATION IMPACTS OF AGRICULTURAL PESTICIDES. Environmental Toxicology and Chemistry 33(7): 1499-1507.

Topping, C. J. and P. Odderskaer (2004): Modeling the influence of temporal and spatial factors on the assessment of impacts of pesticides on skylarks. Environmental Toxicology and Chemistry 23(2): 509-520.

Topping, C. J., P. Odderskaer and J. Kahlert (2013): Modelling Skylarks (Alauda arvensis) to Predict Impacts of Changes in Land Management and Policy: Development and Testing of an Agent-Based Model. Plos One 8(6): e65803.

Topping, C. J., R. M. Sibly, H. R. Akcakaya, G. C. Smith and D. R. Crocker (2005): Risk assessment of UK skylark populations using life-history and individual-based landscape models. Ecotoxicology 14(8): 925-936.

Traas, T. P. and T. Aldenberg (1992): CATS-1: A Model for predicting contaminant accumulation in a meadow ecosystem. The case of cadmium. National Institute of Public Helath and Environmental Protection; Bilthoven, The Netherlands. Document No. 719103001.

Traas, T. P. and T. Aldenberg (1996): Application of CATS models for regional risk-assessment of toxicants. National Institute of Public Health and the Environment; Bilthoven, Netherlands. Report No. 719102039.

Traas, T. P., J. H. Hanse, T. Aldenberg and T. C. M. Brock (1998): A food web model for fate and direct and indirect effects of Dursban(r) 4E (active ingredient chlorpyrifos) in freshwater microcosms. Aquatic Ecology 32: 179-190.

Traas, T. P., J. H. Janse, P. J. Van den Brink, T. C. M. Brock and T. Aldenberg (2004): A freshwater food web model for the combined effects of nutrients and insecticide stress and subsequent recovery. Environmental Toxicology and Chemistry 23(2): 521-529.

Traas, T. P., J. A. Stab, P. R. G. Kramer, W. P. Cofino and T. Aldenberg (1996): Modeling and risk assessment of tributyltin accumulation in the food web of a shallow freshwater lake. Environmental Science & Technology 30(4): 1227-1237.

Trapp, S. (2000): Modelling uptake into roots and subsequent translocation of neutral and ionisable organic compounds. Pest Management Science 56(9): 767-778.

Trapp, S. and T. Eggen (2013): Simulation of the plant uptake of organophosphates and other emerging pollutants for greenhouse experiments and field conditions. Environmental Science and Pollution Research 20(6): 4018-4029.

Trapp, S., M. Matthies, I. Scheunert and E. M. Topp (1990): Modeling the bioconcentration of organic chemicals in plants. Environmental Science & Technology 24(8): 1246-1252.

Trapp, S., C. McFarlane and M. Matthies (1994): Model for uptake of xenobiotics into plants: Validation with bromacil experiments. Environmental Toxicology and Chemistry 13(3): 413-422.

Trapp, S. and J. C. McFarlane (1995). Plant contamination. Modeling and simulation of organic chemical processes. Lewis Publishers; Boca Raton, FL, USA.

Truszkowski, J. (1982): The Impact of the Common Vole on the Vegetation of Agroecosystems. Acta Theriologica 27(23): 305-345.

UBA (2010): Assessment Report on Deltamethrin. Federal Environment Agency Germany, UBA; Report No. ZAA 042973-00/01.

Urrestarazu Ramos, E., W. H. J. Vaes, P. Mayer and J. L. M. Hermens (1999): Algal growth inhibition of *Chlorella pyrenoidosa* by polar narcotic pollutants: toxic cell concentrations and QSAR modeling. Aquatic Toxicology 46(1): 1-10.

USEPA (2000): Aquatox for Windows - Volume 3: Model Validation Reports. Report No. EPA-823-R-00-008.

USEPA (2001): AQUATOX for Windows - Volume 3: Model Validation Reports. Addendum: Formulation, Calibration and Validation of a Periphyton Submodel for AQUATOX Release 1.1. United States Environmental Protection Agency (USEPA). Office of Water. Office of Science and Technology; Washington DC 20460, USA. Report No. EPA-823-R-01-008.

USEPA (2006): Approaches for the application of Physiologically Based Pharmacokinetic (PBPK) models and supporting data in risk assessment. U.S. Environmental Protection Agency; Washington DC, USA. Report No. EPA/600/R-05/043F. Document No. EPA/600/R-05/043F.

USEPA (2013): Sensitivity Analysis of AQUATOX. Office of Water (4305); Document No. EPA- 820-R-13-011.

USEPA (2018): Linking Water Quality and Aquatic Life with AQUATOX 3.2 - Relase 3.2 Fact Sheet. Document No. EPA/600/F-18/329.

Van de Meutter, F., R. Stoks and L. De Meester (2005): Spatial avoidance of littoral and pelagic invertebrate predators by Daphnia. Oecologia 142(3): 489-499.

van den Brink, N. W., D. R. Lammertsma, W. J. Dimmers and M. C. Boerwinkel (2011a): Cadmium Accumulation in Small Mammals: Species Traits, Soil Properties, and Spatial Habitat Use. Environmental Science & Technology 45(17): 7497-7502.

Van den Brink, P. J. and J. M. Baveco (2009). MASTEP, an individual based model to predict recovery of aquatic invertebrates following pesticide stress. Book chapter in: Ecological Models for Regulatory Risk Assessments of Pesticides: Developing a strategy for the future. P. Thorbek, V. E. Forbes, F. Heimbach et al. SETAC and CRC Press.

Van den Brink, P. J., J. M. Baveco, J. Verboom and F. Heimbach (2007a): An individual-based approach to model spatial population dynamics of invertebrates in aquatic ecosystems after pesticide contamination. Environmental Toxicology and Chemistry 26(10): 2226-2236.

Van den Brink, P. J., C. D. Brown and I. G. Dubus (2006): Using the expert model PERPEST to translate measured and predicted pesticide exposure data into ecological risks. Ecological Modelling 191(1): 106-117.

Van den Brink, P. J., J. Roelsma, E. H. Van Nes, M. Scheffer and T. C. M. Brock (2002): PERPEST model, a case-based reasoning approach to predict ecological risks of pesticides. Environmental Toxicology and Chemistry 21(11): 2500-2506.

Van den Brink, P. J., I. Roessink, P. I. Adriaanse, W. H. J. Beltman, J. Verboom, P. Schippers and H. Baveco (2007b): Landscape Simulation of Temporal and Spatial Effects on Asellus aquaticus from Concentrations of Pyridalyl in Surface Water. Alterra Wageningen University and Research Centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0058. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome, Chuo-ku, Tokyo 104-8260, Japan.

Van den Brink, P. J., I. Roessink, P. I. Adriaanse, W. H. J. Beltman, J. Verboom, P. Schippers and J. M. Baveco (2011b): Annex 1 to Landscape Simulation of Temporal and Spatial Effects on Asellus aquaticus from Concentrations of Pyridalyl in Surface Water. Alterra Wageningen University and Research Centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0058. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome, Chuo-ku, Tokyo 104-8260, Japan.

Van den Brink, P. J., I. Roessink, P. I. Adriaanse, W. H. J. Beltman, J. Verboom, P. Schippers and J. M. Baveco (2011c): Annex 2 to Landscape Simulation of Temporal and Spatial Effects on Asellusaquaticus from Concentrations of Pyridalyl in Surface Water. Alterra Wageningen University and Research Centre; PO Box 47, 6700 AA Wageningen, The Netherlands. Report No. SUW-0058. Sponsor: Sumitomo Chemical Company Ltd., 27-1 Shinkawa 2-chome, Chuo-ku, Tokyo 104-8260, Japan.

Van der Vaart, E., A. S. A. Johnston and R. M. Sibly (2016): Predicting how many animals will be where: how to build, calibrate and evaluate individual-based models. Ecological Modelling 326: 113-123.

Van Eijkeren, J. C. H., M. J. Zeilmaker, C. A. Kan, W. A. Traag and L. A. P. Hoogenboom (2006): A toxicokinetic model for the carry-over of dioxins and PCBs from feed and soil to eggs. Food Additives and Contaminants 23(5): 509-517.

Van Haren, R. J. F., H. E. Schepers and S. A. L. M. Kooijman (1994): Dynamic Energy Budgets affect kinetics of xenobiotics in the marine mussel *Mytilus edulis*. Chemosphere 29(2): 163-189.

van Heesch, D. (1997). http://www.doxygen.nl/10.04.2019.

Van Nes, E. H. and P. J. Van den Brink (2003): PERPEST version 1.0, manual and technical description - A model that Predicts the Ecological Risks of PESTicides in freshwater ecosystems. Alterra rapport. G. W. R. Alterra. Wageningen, Alterra, Green World Research. 787.

van Wijngaarden, R. P. A., G. H. P. Arts, J. D. M. Belgers, H. Boonstra, I. Roessink, A. F. W. Schroer and T. C. M. Brock (2010): The species sensitivity distribution approach compared to a microcosm study: A case study with the fungicide fluazinam. Ecotoxicology and Environmental Safety 73(2): 109-122.

Vanoverbeke, J. (2008): Modeling individual and population dynamics in a consumer-resource system: Behavior under food limitation and crowding and the effect on population cycling in Daphnia. Ecological Modelling 216(3-4): 385-401.

Verboom, J., H. Baveco and P. J. Van den Brink (2005): A Simulation Model for Spatial Population Dynamics of Asellus aquaticus after a Spray Drift Event of Deltamethrin in Aquatic Ecosystems. Alterra Wageningen University and Research Center; PO Box 47, 6700 AA Wageningen, The Netherlands. Sponsor: Bayer Crop Science AG, Development - Ecotoxicology, 40789 Monheim, Germany.

Von Der Ohe, P. C. and M. Liess (2004): Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. Environmental Toxicology and Chemistry 23(1): 150-156.

Waage, J. K., M. P. Hassell and H. C. J. Godfray (1985): The Dynamics of Pest Parasitoid Insecticide Interactions. Journal of Applied Ecology 22(3): 825-838.

Wang, M. (2013): From home range dynamics to population cycles: Validation and realism of a common vole population model for pesticide risk assessment. Integrated Environmental Assessment and Management 9(2): 294-307.

Wang, M. and V. Grimm (2007): Home range dynamics and population regulation: An individual-based model of the common shrew Sorex ayaneus. Ecological Modelling 205(3-4): 397-409.

Wang, M. and V. Grimm (2010): Population models in pesticide risk assessment: Lessons for assessing population-level effects, recovery, and alternative exposure scenarios from modeling a small mammal. Environmental Toxicology and Chemistry 29(6): 1292-1300.

Widarto, T. H., M. Holmstrup and V. E. Forbes (2004): The influence of nonylphenol on life-history of the earthworm *Dendrobaena octaedra* Savigny: linking effects from the individual- to the population-level. Ecotoxicology and Environmental Safety 58(2): 147-159.

Widmark, E. and J. Tandberg (1924): Über die Bedingungen für die Akkumulation indifferenter Narkotika. Theoretische Berechnungen. Biochemische Zeitschrift 147: 358-369.

Wiertz, S. (1984). Funktionelle Reaktion bei den Larven der Mücke Chaoborus crystallinus (Diptera, Chaoboridae) unter besonderer Berücksichtigung der Temperatur- und Photoperiodeabhängigkeit. Diploma Thesis Thesis, RWTH Aachen University.

Wissinger, S., J. Steinmetz, J. S. Alexander and W. Brown (2004): Larval cannibalism, time constraints, and adult fitness in caddisflies that inhabit temporary wetlands. Oecologia 138(1): 39-47.

Wogram, J. and M. Liess (2001): Rank ordering of macroinvertebrate species sensitivity to toxic compounds by comparison with that of daphnia magna. Bull Environ Contam Toxicol 67: 360-367.

Wolton, R. J. (1985): The ranging and nesting behavior of wood mice, Apodemus sylvaticus(Rodentia: Muridae), as revealed by radio-tracking. Journal of Zoology 206: 203–224.

Wu, Y., S. M. Bartell, J. Orr, J. Ragland and D. Anderson (2010): A risk-based decision model and risk assessment of invasive mussels. Ecological Complexity 7(2): 243-255.

Zhang, L., J. Liu, Y. Li and Y. Zhao (2013): Applying AQUATOX in determining the ecological risk assessment of polychlorinated biphenyl contamination in Baiyangdian Lake, North China. Ecological Modelling 265: 239-249.

Zhao, J. and B. Chen (2016): Species sensitivity distribution for chlorpyrifos to aquatic organisms: Model choice and sample size. Ecotoxicology and Environmental Safety 125: 161-169.

Zimmer, E. I. (2017): Rainbow Trout: Dynamic Energy Budget (DEB) model simulation to extrapolate effects of beta-cyfluthrin on growth from constant exposure to time-varying exposure. ibacon GmbH; Arheiliger Weg 17, 64380 Rossdorf, Germany. Report No. 111871520. Sponsor: Bayer AG – CropScience Division, Effect Modelling, 40789 Monheim Am Rhein, Germany. Sponsor Ref. No. MOAA0004.

Zimmer, E. I., T. Jager, V. Ducrot, L. Lagadic and S. A. L. M. Kooijman (2012): Juvenile food limitation in standardized tests: a warning to ecotoxicologists. Ecotoxicology 21: 2195-2204.

Zimmer, E. I., T. G. Preuss, S. Norman, B. Minten and V. Ducrot (2018): Modelling effects of time-variable exposure to the pyrethroid beta-cyfluthrin on rainbow trout early life stages. Environmental Sciences Europe 30(1): 36.

Zonneveld, C. and S. A. L. M. Kooijman (1993): Comparative kinetics of embryo development. Bulletin of Mathematical Biology 55(3): 609-635.