TEXTE

41/2024

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

Annex

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
FB001319/ENG

Annex

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

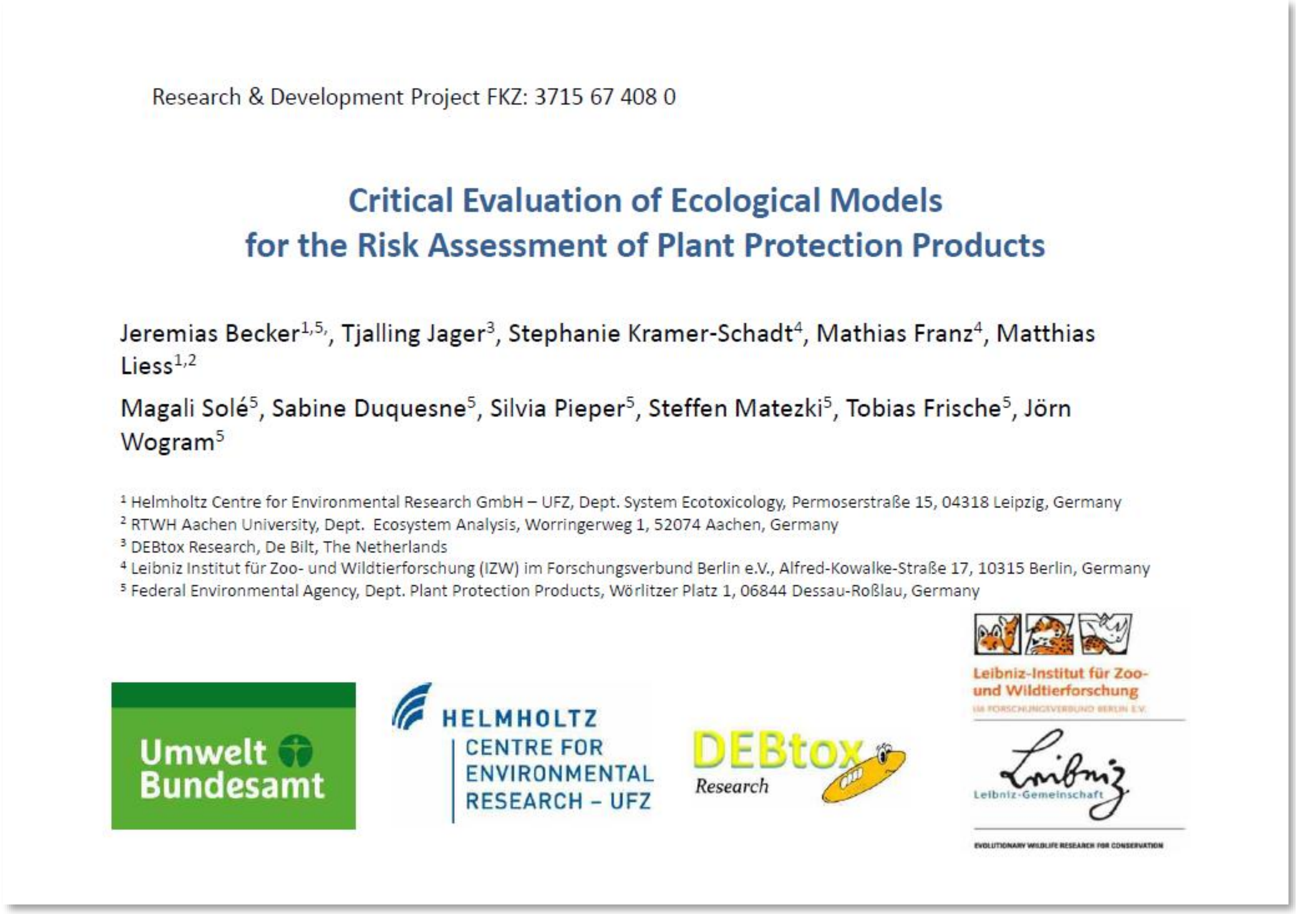
Dessau-Roßlau, March 2024

Table of Contents

List of Figures

Annex 1: Presentations at the Symposium

Presentations held by the members of the consortium at the symposium ”Evaluation of toxicological and ecological effect models for risk assessment of plant protection products”, organized by UFZ and UBA in Berlin, 19th and 20th of September 2019, as part of the UFO Plan project 3715 67 408 0.



Presentation on TKTD Models

Individual-level effects GUTS

Tjalling Jager

Presentation at UBA Symposium
19 Sept. 2019

DEBtox Research

Contents

Individual-level effects models

- Introduction TKTD/GUTS
- Notes on evaluation of GUTS
- Evaluation of case study

Discussion

- Introduction DEBtox
- Evaluation of case study

Discussion

Individual-level models

- Static models (e.g., dose-response curves)

external concentration → **Static model** → *effects at one arbitrary time point*

- ToxicoKinetic (TK) models

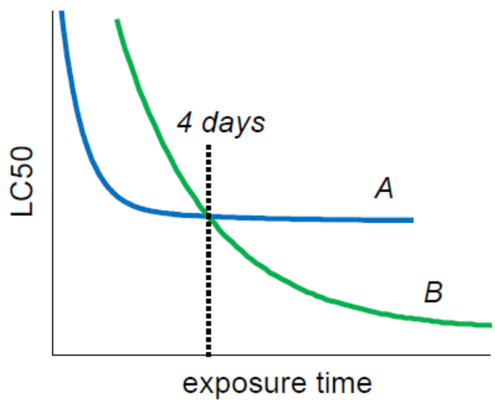
external concentration → **TK model** → *internal concentration over time*

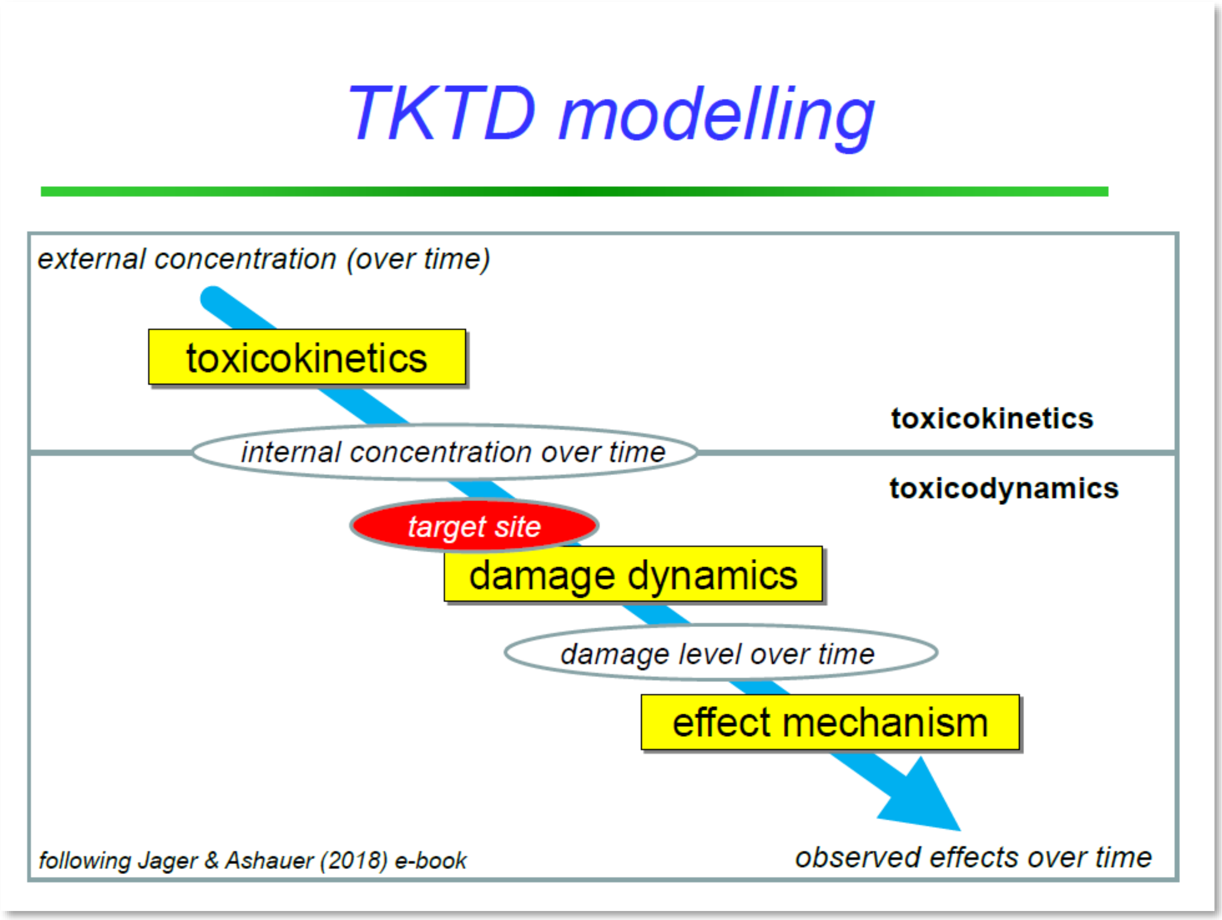
- ToxicoKinetic-ToxicoDynamic (TKTD) models

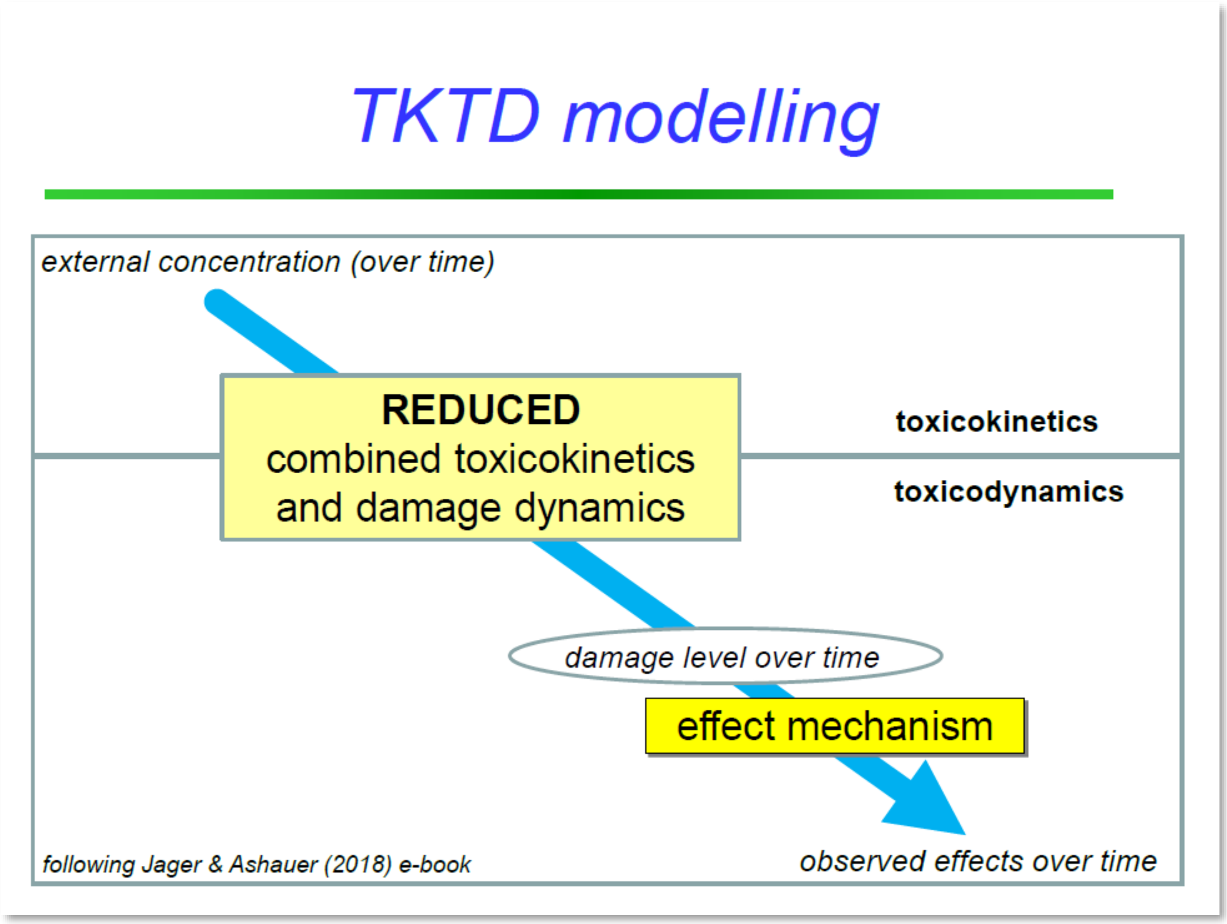
external concentration → **TK model** → **TD model** → *effects over time*

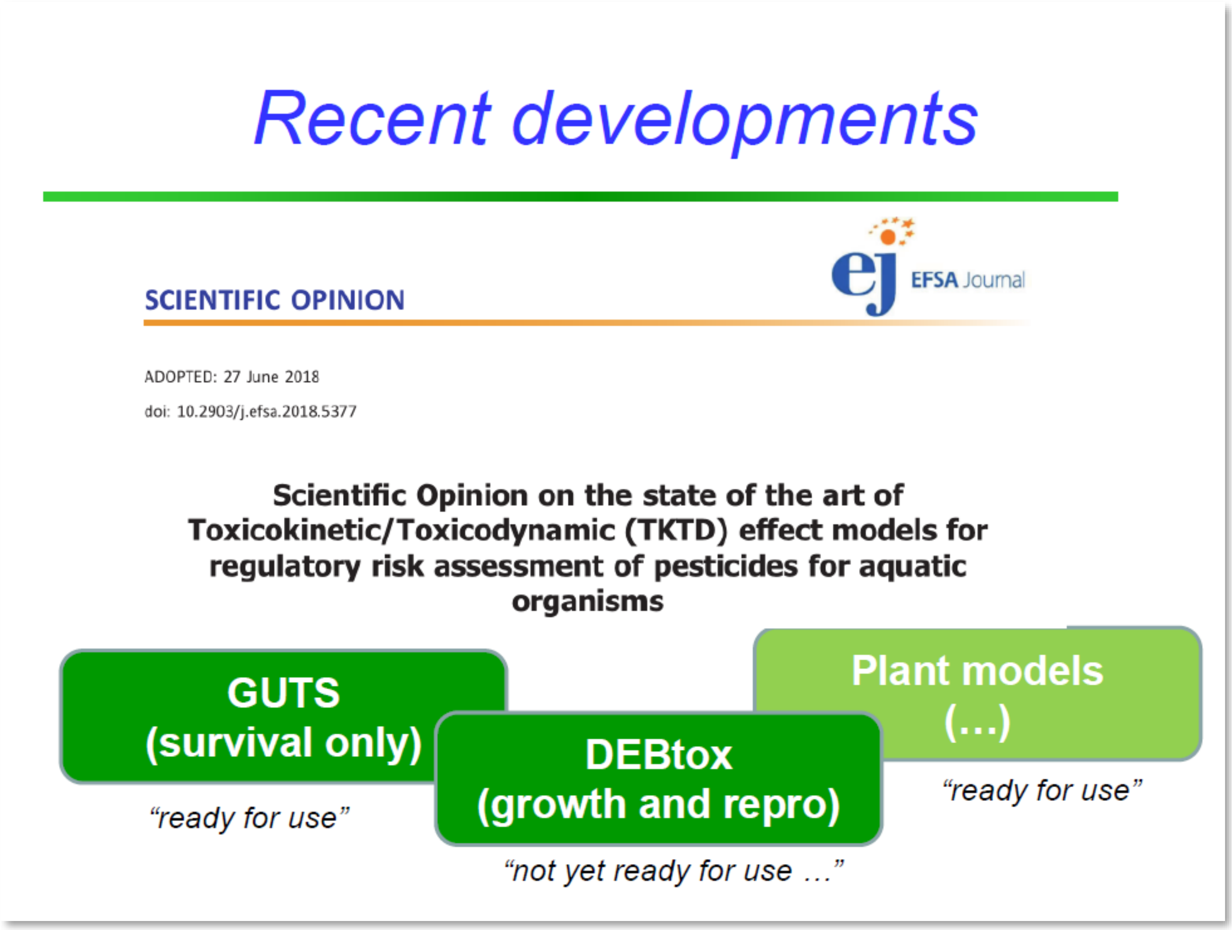
TKTD modelling

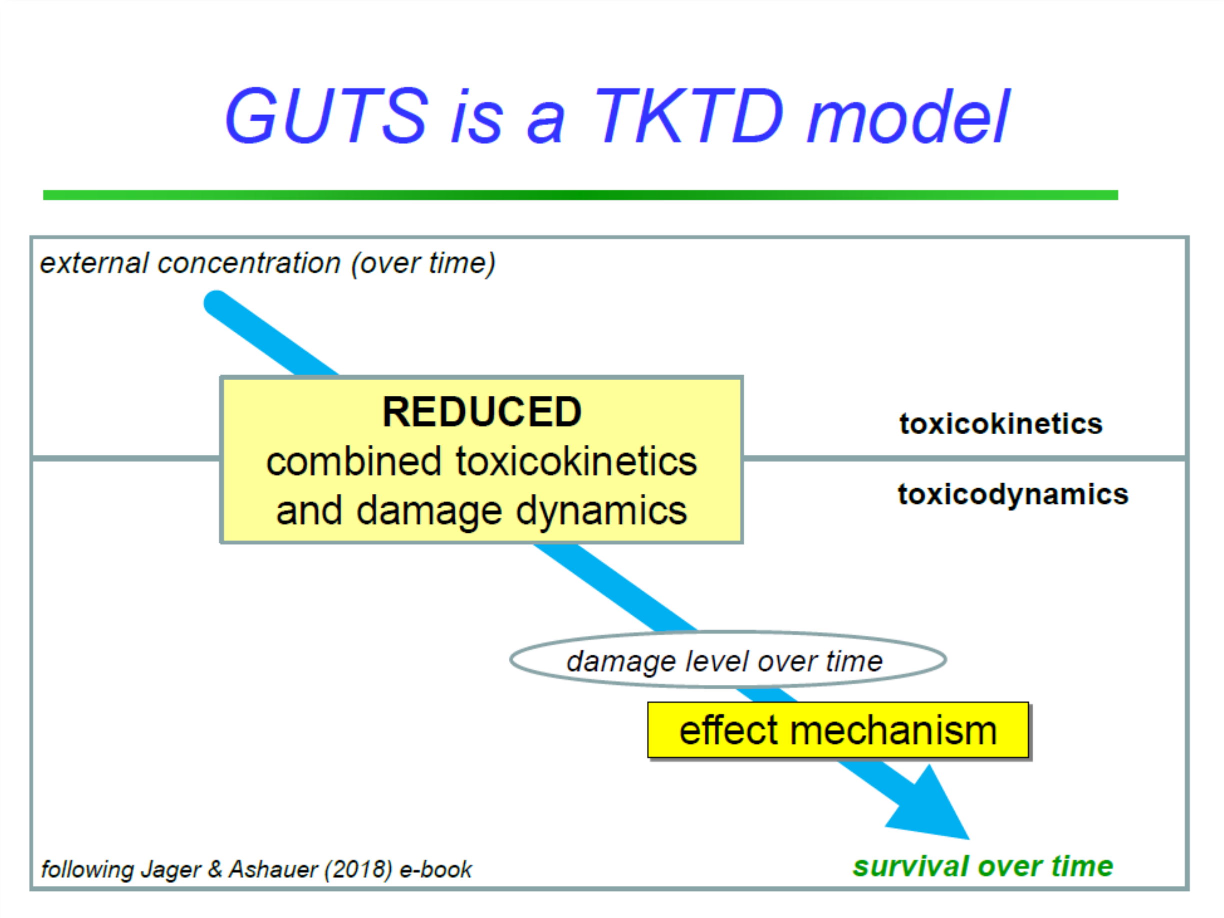
- Aim of TKTD models:
  - replace static models for dose-response analysis
  - make predictions for untested exposure profiles
  - building block for population/community models
- What's wrong with static?
  - poor use of data
  - ECx depends on time
  - no meaningful extrapolation (e.g., to time-var. exposure)
  - inconsistent conservatism




TKTD modelling
external concentration (over time)
toxicokinetics
internal concentration over time
toxicokinetics
toxicodynamics
target site
damage dynamics
damage level over time
effect mechanism
following Jager & Ashauer (2018) e-book
observed effects over time


TKTD modelling
external concentration (over time)
REDUCED
combined toxicokinetics
and damage dynamics
toxicokinetics
toxicodynamics
damage level over time
effect mechanism
following Jager & Ashauer (2018) e-book
observed effects over time

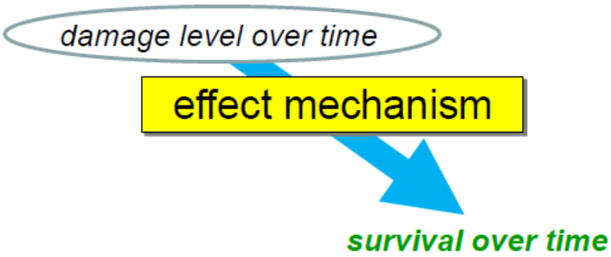

Recent developments
SCIENTIFIC OPINION
EFSA Journal
ADOPTED: 27 June 2018
doi: 10.2903/j.efsa.2018.5377
Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms
GUTS (survival only)
"ready for use"
DEBtox (growth and repro)
"not yet ready for use …"
Plant models (…)
"ready for use"


GUTS is a TKTD model
external concentration (over time)
REDUCED combined toxicokinetics and damage dynamics
toxicokinetics
toxicodynamics
damage level over time
effect mechanism
survival over time
following Jager & Ashauer (2018) e-book

GUTS is a TKTD model

Effect mechanism

- death represented as a chance process
- simplest cases: just 3-4 model parameters (all fitted)

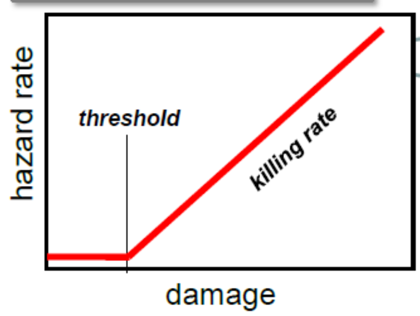


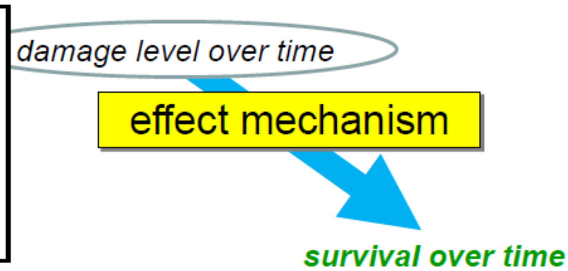
GUTS is a TKTD model

Effect mechanism

- death represented as a chance process
- simplest cases: just 3-4 model parameters (all fitted)

stochastic death (SD)
chance: individual
damage>threshold
increases probability to
die





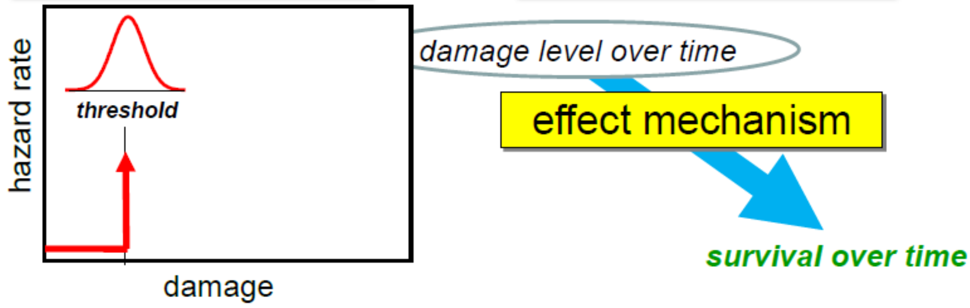
GUTS is a TKTD model

Effect mechanism

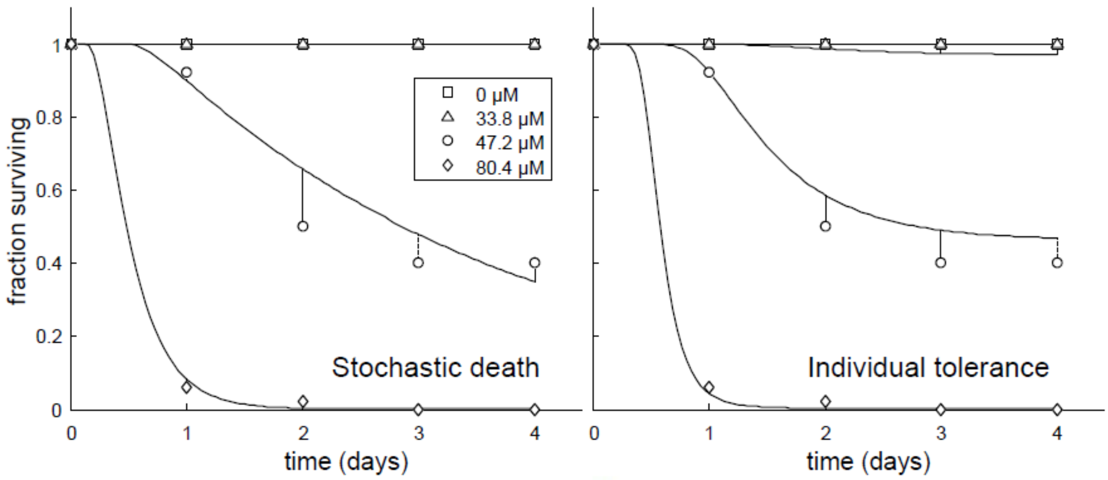
- death represented as a chance process
- simplest cases: just 3-4 model parameters (all fitted)

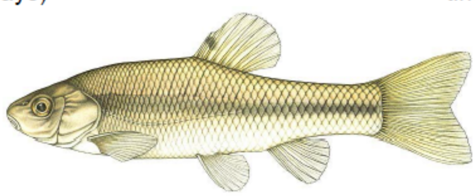
stochastic death (SD)
chance: individual
damage>threshold increases probability to die

individual tolerance (IT)
chance: cohort
individual draws threshold value from a distribution



Example naphthalene

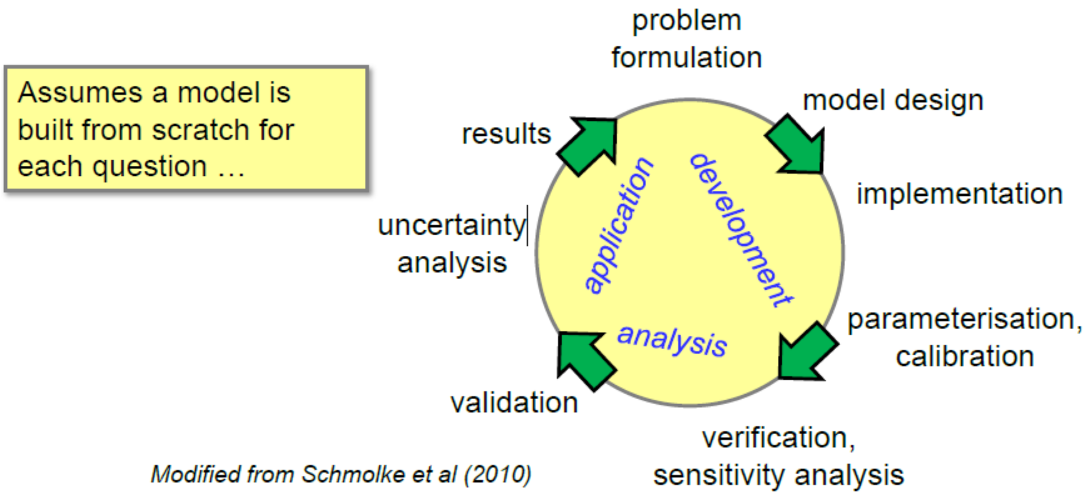




Notes on model evaluation

'Good modelling practice'

- TRACE (Schmolke *et al* 2010, Grimm *et al* 2014)
- EFSA scientific opinion (2014)

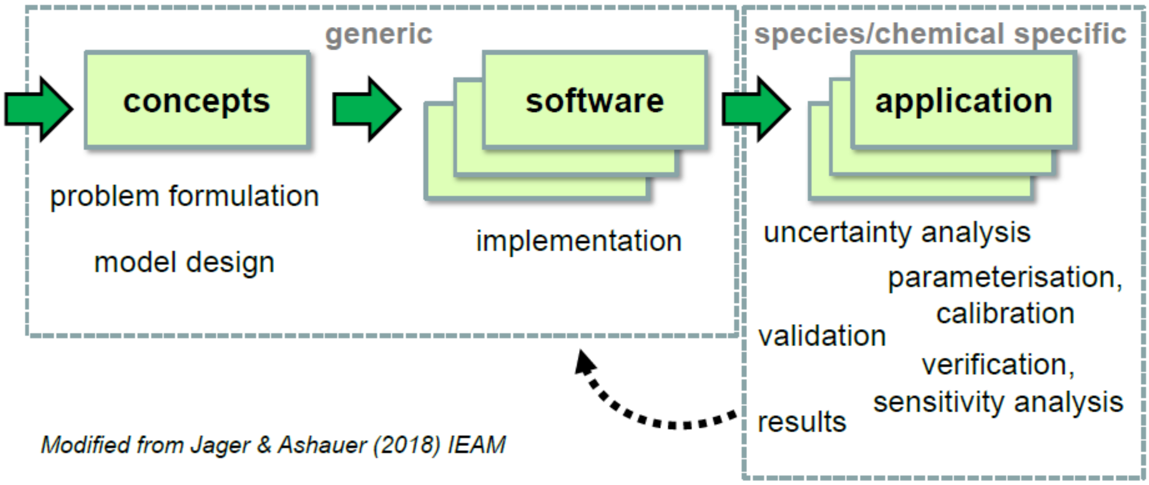


Modified from Schmolke et al (2010)

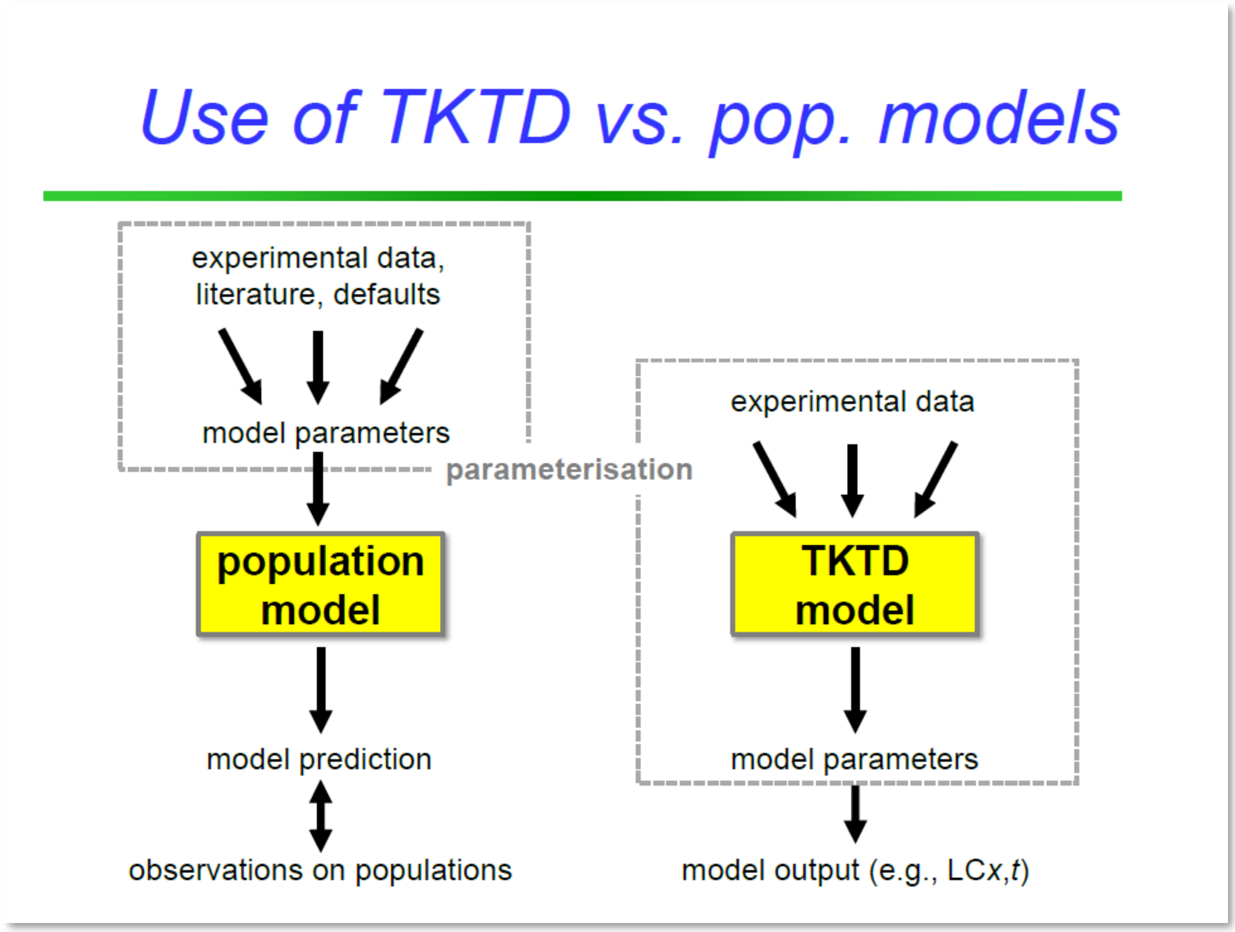
Notes on model evaluation

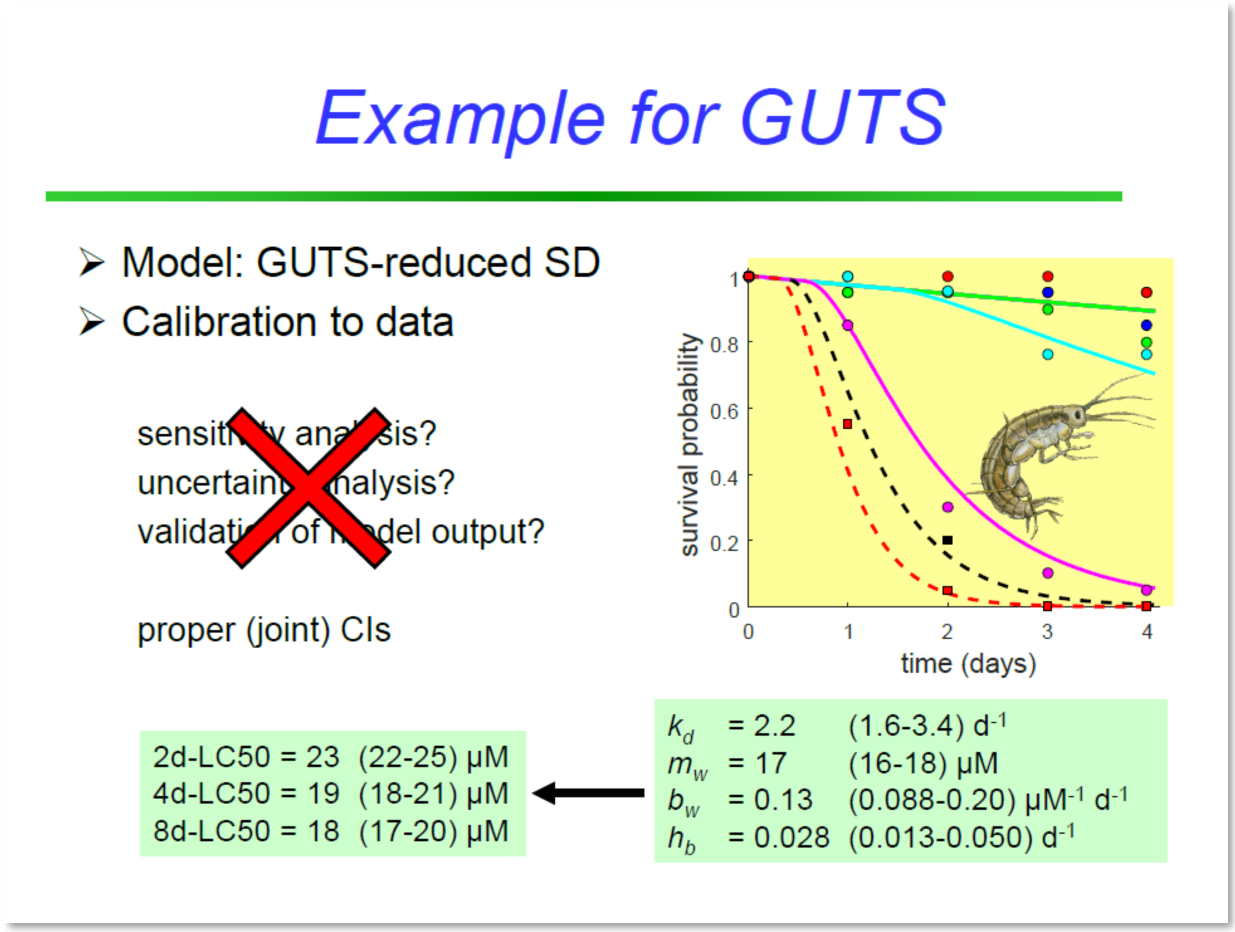
'Good modelling practice'

- TRACE (Schmolke *et al* 2010, Grimm *et al* 2014)
- EFSA scientific opinion (2014)



Modified from Jager & Ashauer (2018) IEAM


Use of TKTD vs. pop. models
experimental data,
literature, defaults
model parameters
parameterisation
population model
model prediction
observations on populations
experimental data
TKTD model
model parameters
model output (e.g., LCx,t)


Example for GUTS
Model: GUTS-reduced SD
Calibration to data
sensitivity analysis?
uncertainty analysis?
validation of model output?
proper (joint) CIs
survival probability
time (days)
0
0.2
0.4
0.6
0.8
1
0
1
2
3
4
2d-LC50 = 23 (22-25) µM
4d-LC50 = 19 (18-21) µM
8d-LC50 = 18 (17-20) µM
kd = 2.2 (1.6-3.4) d-1
mw = 17 (16-18) µM
bw = 0.13 (0.088-0.20) µM-1 d-1
hb = 0.028 (0.013-0.050) d-1

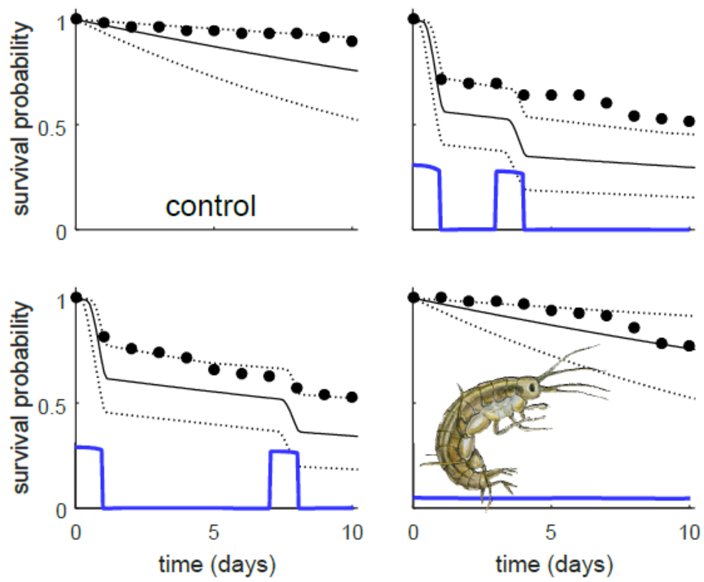
Predictions allow 'validation'

- Calibrated model to predict data *not* used for calibration

validation only possible for specific chemical+species

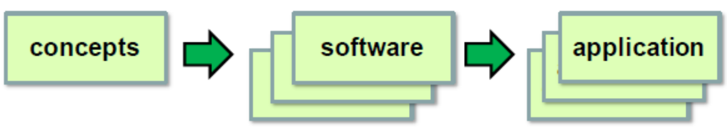
~~sensitivity/uncertainty analysis?~~

error propagation



General evaluation of GUTS

- Simple modelling cycle is misleading (esp. for TKTD)
- Many items on EFSA GMP checklist don't apply:
  - evaluation of supporting data
  - evaluation of sensitivity and uncertainty analysis
  - comparison with data from independent measurements
  - …
- Current approaches for effects in ERA are also models
  - makes sense to evaluate them in the same manner



Concepts

- GUTS unifies all existing TKTD models for survival
  - therefore, there is simply no alternative left
  - builds on >100 years history in survival analysis
- Status of GUTS (and predecessors)
  - large and diverse user community
  - many publications (±100 on SD models alone, since 1994)
  - broad acceptance, EFSA (2018): GUTS is “ready for use”

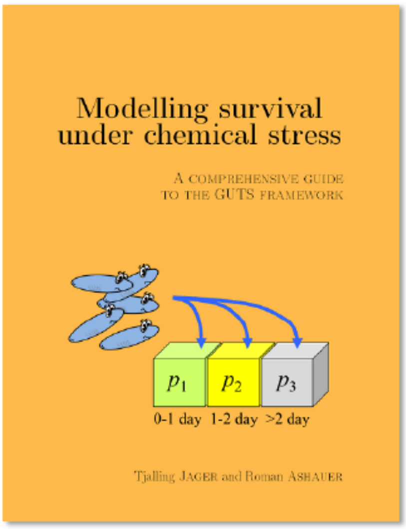
SCIENTIFIC OPINION   EFSA Journal

ADOPTED: 27 June 2018
doi: 10.2903/j.efsa.2018.5377

Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms

Documentation

- Extensive (free) e-book
- Contains a.o.:
  - conceptual description
  - mathematical description
  - statistical background
  - case studies
  - ring-test of 11 implementations



Software

- Range of software implementations
  - each with their advantages and limitations
  - standalone and platform-based (Matlab, R etc.)
  - Bayesian and 'frequentist'
  - differences in user-friendliness, features, etc.
- 2017 Ringtest: Jager & Ashauer (2018), e-book

MOSAIC openGUTS

Suitability for RA

EFSA (2018): "ready for use"

- Advantages
  - optimal use of animal data (std. and non-std.)
  - identifies problems with chemical (e.g., irreversible effects)
  - robust LC*x*,*t* for any *x* and any *t*
  - meaningful evaluation of exposure profiles (LP*x*)
  - greater consistency in degree of conservatism
  - future-proof (e.g., mixtures, comparing species/chemicals)
- Model is huge simplification and thus 'wrong'
  - but, not as wrong as using peak conc. and 4d-LC50 …
  - note: GUTS does not necessarily yield lower risk …

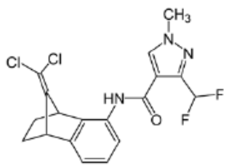
Suitability for RA

General uncertainties/limitations

- Only predict effects from mechanisms seen in tests
  - non-tested life stages/species may be specifically sensitive
  - long-term, low-intensity, exposure may reveal other mechanisms
  - acute tests carry only limited information …
- Does not account for
  - different conditions in the field (e.g., other stressors)
  - growth/development of individuals
- All extrapolations rest on assumptions that:
  - model is true, parameters remain constant
- Judging fit/validation requires expertise/experience

GUTS case study

- Benzovindiflupyr
  - broad-spectrum pyrazole carboxamide fungicide
  - critical issues for acute risk to fish
- Published in open literature (Ashauer *et al.*, 2013)



SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 32, No. 4, pp. 954–965, 2013
© 2013 SETAC
Printed in the USA
DOI: 10.1002/etc.2144

A METHOD TO PREDICT AND UNDERSTAND FISH SURVIVAL UNDER DYNAMIC CHEMICAL STRESS USING STANDARD ECOTOXICITY DATA

Roman Ashauer,*†‡ Pernille Thorbek,§ Jacqui S. Warinton,§ James R. Wheeler,§ and Steve Maund||
†Environment Department, University of York, York, United Kingdom,
‡Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
§Syngenta, Environmental Safety, Berkshire, United Kingdom,
||Syngenta Crop Protection AG, Basel, Switzerland

Problem definition

Field exposure highly time-varying; toxicity data for 4-d constant exposure

- Aim:
  - extrapolate 4-day tox test to FOCUS-SW profiles
  - calculate safety margin for 10% effect (LP10)
- Analysis follows workflow adopted (later) by EFSA:
  1. calibrate model to (standard) data
  2. validate with independent data
  3. predict survival for FOCUS profile (derive safety margin)

  both for SD and IT

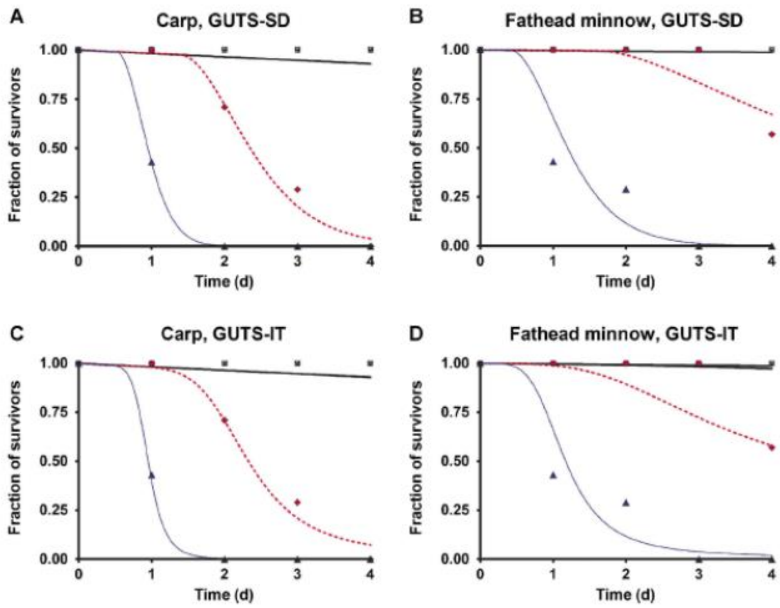
Calibration: supporting data

Calibration

- 4-day acute toxicity tests carp and fathead minnow
  - most sensitive of 5 species (also under FOCUS scenarios!)
  - fit for purpose, but few individuals per treatment (7) and few observations (4)
  - background hazard fixed to control performance

Parameter estimation

- Fits look good enough (with n=7)



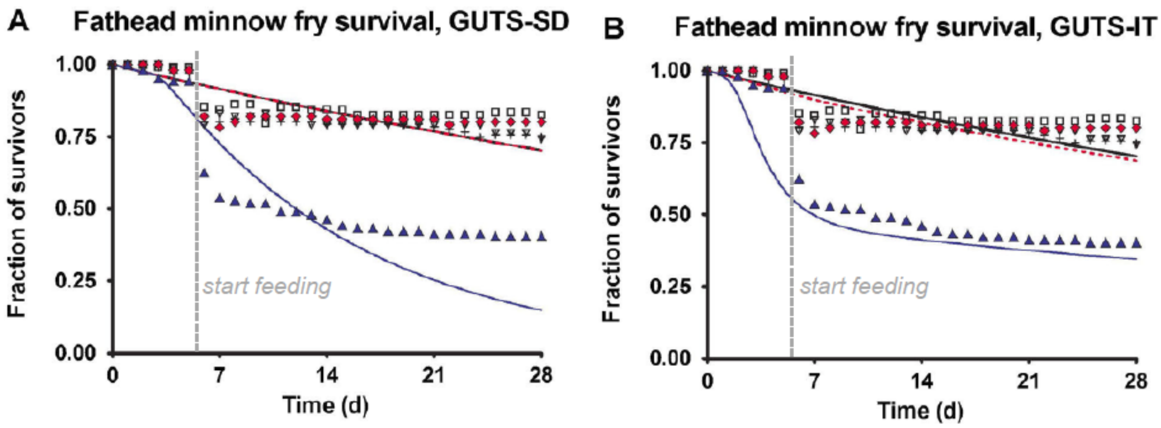
Validation: supporting data

ELS study with fathead minnow

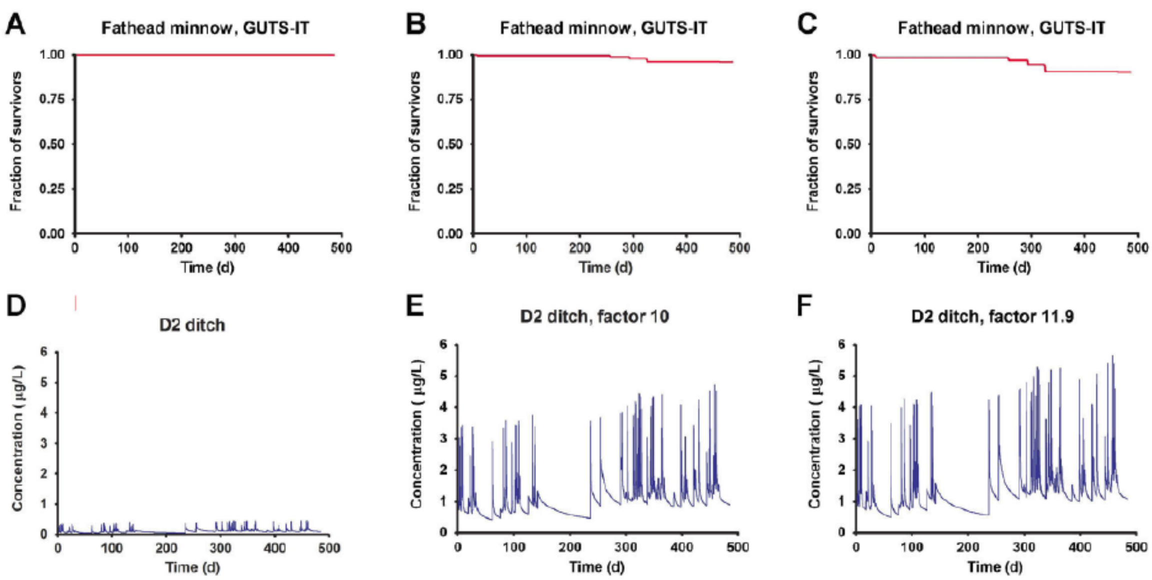
- Little support for validity of calibrated model
  - different life stage, animal are fed and growing/developing …
- Cannot support extrapolation to time-varying exposure
  - constant exposure, just like calibration test
- Not consistent with requirements EFSA SO
  - but, for vertebrates, case-by-case evaluation …

Validation: comparison

- Validation with ELS study
  - supports distinguishing toxic and non-toxic concentrations
  - shows that early stages are not more sensitive
  - some support extrapolation to longer durations



Prediction (LP10)

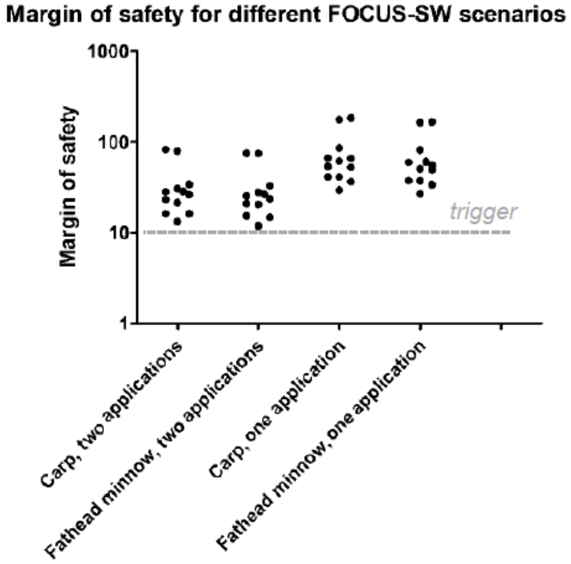


Prediction (LP10)

- LP10 values range: 26-500 (DAR) or 12-184 (paper)

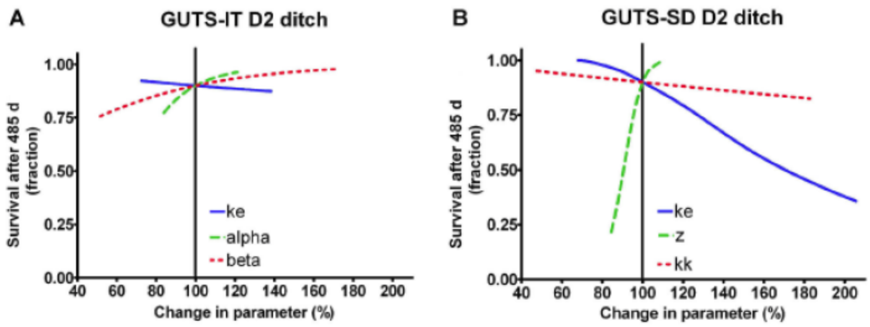
points are 6 FOCUS scenarios, for SD and IT

10x safety margin acceptable for RMS:
- 5 species tested
- for 10% mortality



Sensitivity/uncertainty

- Calibration: method for parameter CIs incorrect
  - coverage will be less than 95%
- Parameter uncertainty not propagated to predictions
  - note: also no CIs on FOCUS profiles …
- Limited sensitivity analysis
  - one parameter at a time, moved over its CI



Conclusions case study

Approach largely in line with EFSA opinion

- Data basis is limited
  - acute tests with 5 species, but only 7 indiv./treatment
  - validation very limited (ELS at constant exposure)
- Some technical problems with analysis
  - incorrect CIs on parameters, no error propagation
- Margin of safety of 10 on LP10 sufficient?
  - EFSA SO (generally) suggests 100, but on LP50
  - what to do with CI?

Questions, comments, discussion

What is DEBtox?

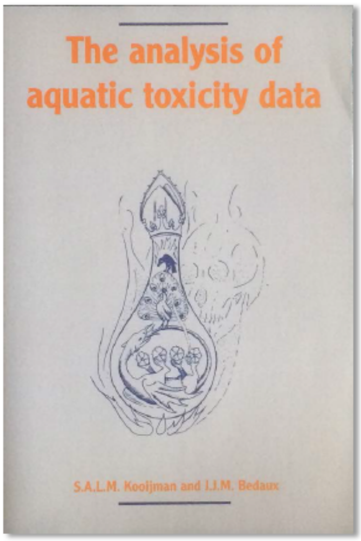
TKTD for growth/development/reproduction

Initially:

- Simplified DEB models:
  - Kooijman & Bedaux (1996)
  - updated by Billoir *et al* (2008) and Jager & Zimmer (2012)

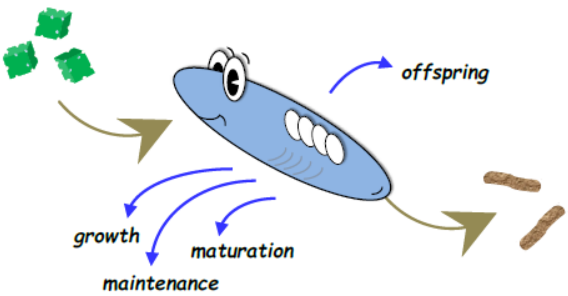
More general:

- *Any* DEB-based model applied to toxicant stress

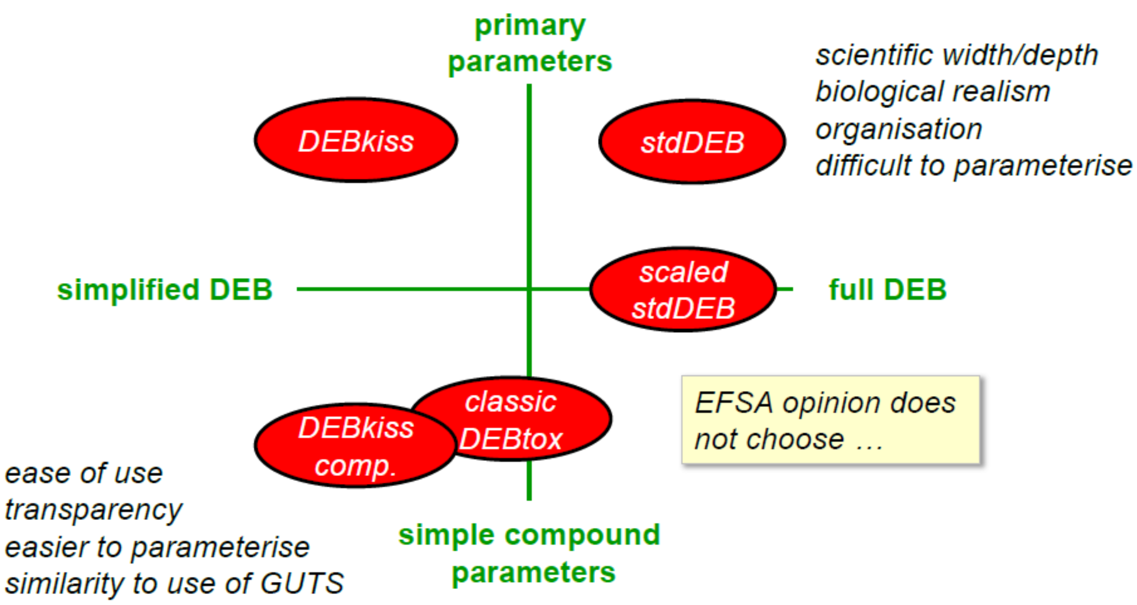


Concepts

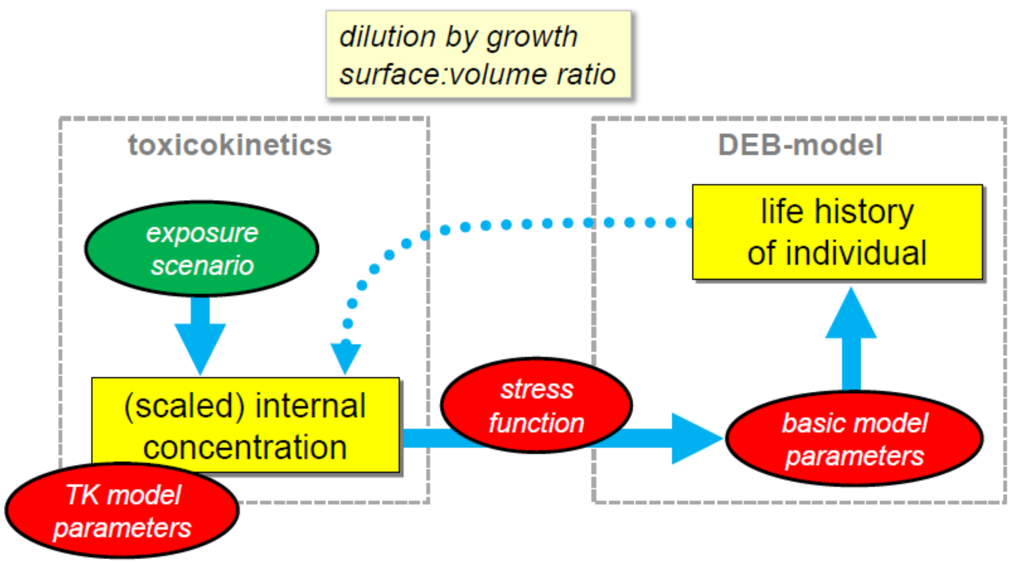
- Based on extensive and well-tested theory
  - only approach for sub-lethal effects with track record
  - approx. 85 publications
- Status of DEBtox
  - included in ISO/OECD guidance (2006)
  - EFSA SO: "not ready for use"
    - lack of relevant case studies
    - lack of user-friendly software



DEB-model family

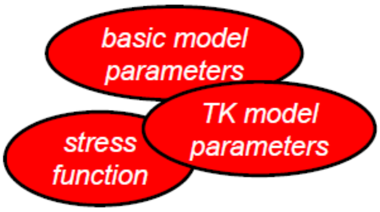


Dealing with toxicants



Strategies for application

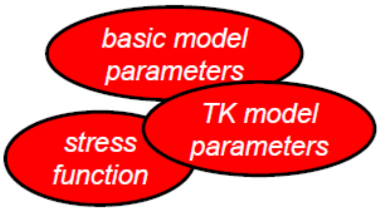
- Simplified DEB models (e.g., 'classic DEBtox')



Fit *all* on toxicity test data only
(growth and reproduction for juveniles-adults)

Strategies for application

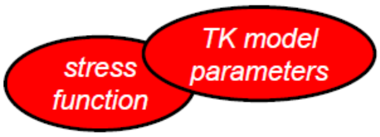
- Simplified DEB models (e.g., 'classic DEBtox')



Fit *all* on toxicity test data only
(growth and reproduction for juveniles-adults)

- Full DEB models (e.g., 'standard DEB')

Take from 'add-my-pet' collection

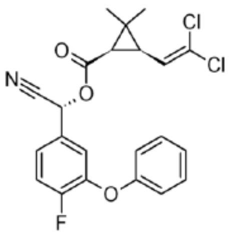


Fit on toxicity test data
(may accommodate more limited test designs)

Case study

Beta-cyfluthrin

- Pyrethroid insecticide
- Published by Zimmer *et al.* (2018)



Zimmer *et al. Environ Sci Eur* (2018) 30:36
https://doi.org/10.1186/s12302-018-0162-0

Environmental Sciences Europe

RESEARCH Open Access

Modelling effects of time-variable exposure to the pyrethroid beta-cyfluthrin on rainbow trout early life stages

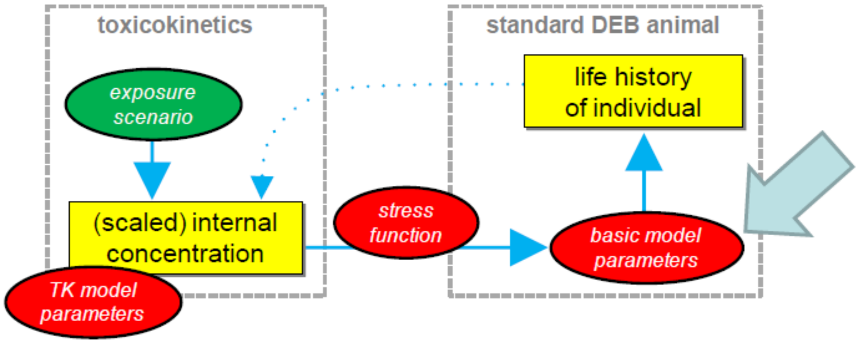
Elke I. Zimmer[1*], Thomas G. Preuss[2], Steve Norman[3], Barbara Minten[4] and Virginie Ducrot[2]

Problem definition

- Model analysis submitted as “additional information”
  - “explain the differences between the observed effects in the two ELS studies”
  - “In the future, … predict the survival and growth … under various realistic peak exposure scenarios ”
- DEBtox generally growth/reproduction juveniles/adults
  - here: ‘standard DEB’ to ELS toxicity
  - this is a novelty (very few relevant case studies)

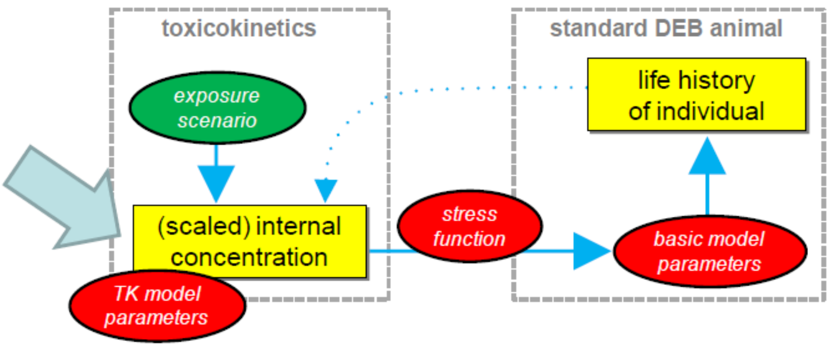
Using standard DEB for ELS

- Basic model parameters from AmP collection
  - >1400 animal species
  - standard format/procedure
  - ‘board of curators’
  - considerable user community
  - not very transparent
  - entries differ in quality/completeness
  - ‘pseudo data’ are added
  - no quantification uncertainties yet
  - life history may differ in test control …



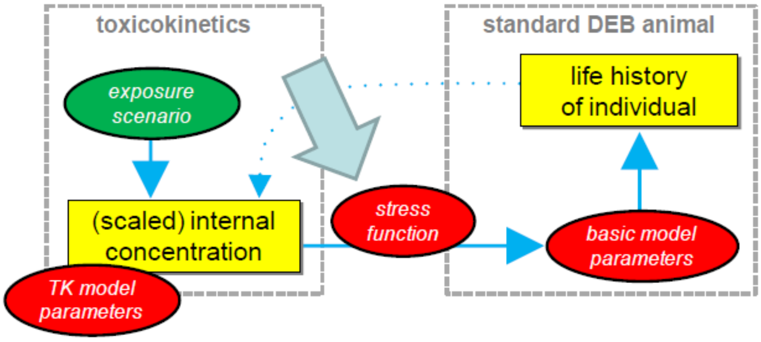
Using standard DEB for ELS

- Standard TK model from classic DEBtox
  - egg/yolk (and reserve) imply specific issues for TK
  - uptake assumed to start with feeding



Using standard DEB for ELS

- Target process 'feeding' assumed
  - only indirect support for this mechanism …



Calibration: supporting data

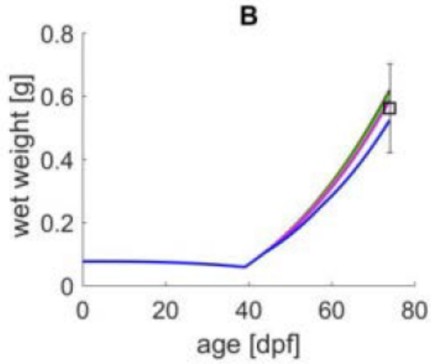
Basic data

- Entry for rainbow trout in add-my-pet collection
  - substantial amount of data for early-life stages
  - note that entry has changed since this study …
- Predicted size does not match data …
  - food level tuned to match control body size at end of test
  - food level roughly 50% of expectation AmP entry
  - discrepancy could have other causes

Calibration: supporting data

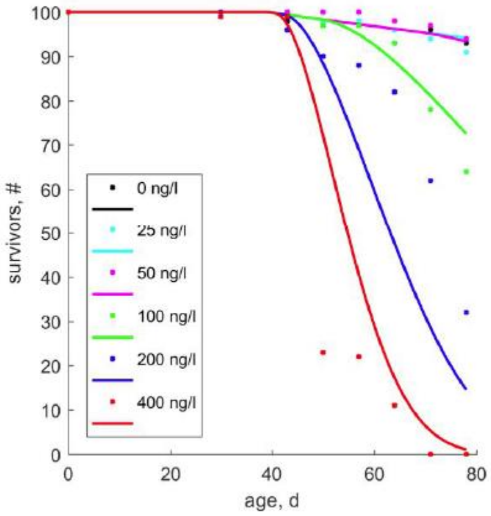
Toxicity data

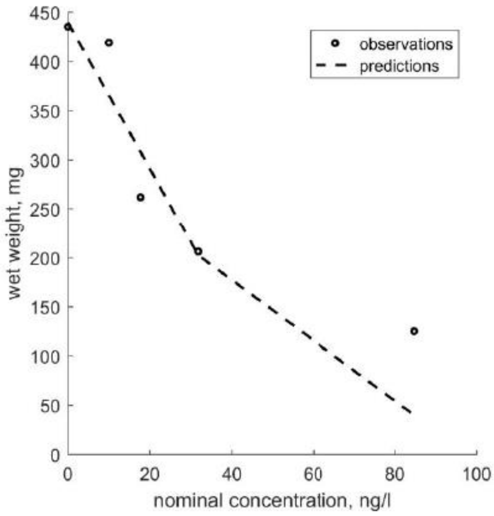
- ELS study with constant exposure
- Insufficient information for TKTD modelling
  - body weight/length only at end of test
  - survival weekly
- Possibly could have used
  - timing of events (hatching, start of feeding)



Parameter estimation

- 7 parameters estimated/tuned on these data
  - uptake assumed to start at start of feeding
  - fits in code are different from those in text





Sensitivity and uncertainty

- No sensitivity/uncertainty analysis performed
- No CIs are provided
  - not on the AmP parameters, not on fitted tox parameters
  - can parameters be uniquely identified from these data?

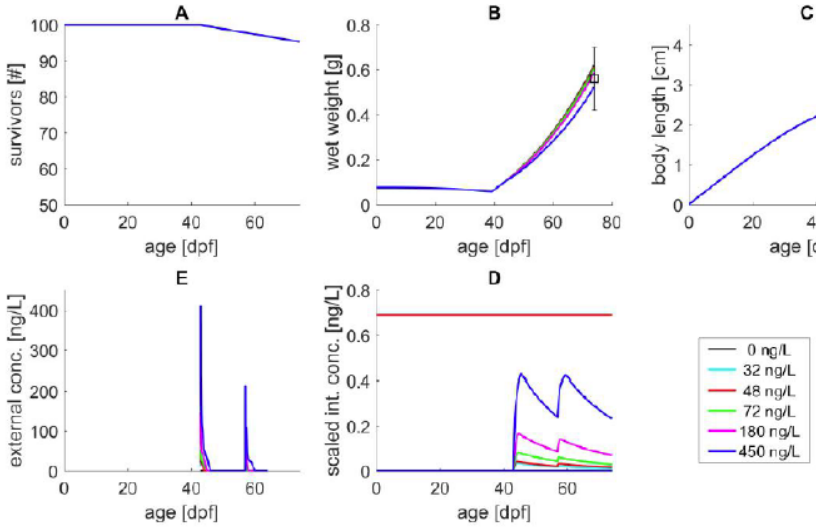
Validation: supporting data

Validation

- ELS study with two pulses
  - three separate tests, with pulses applied in different life stages (egg, sac-fry and swim-up fry)
- Issues with data
  - validation data set showed hardly any effects …
  - possible more info (feeding behaviour, timing of events)
  - different test conditions (addition of sediment)
  - food level tuned to match final body size (all three cohorts independently)

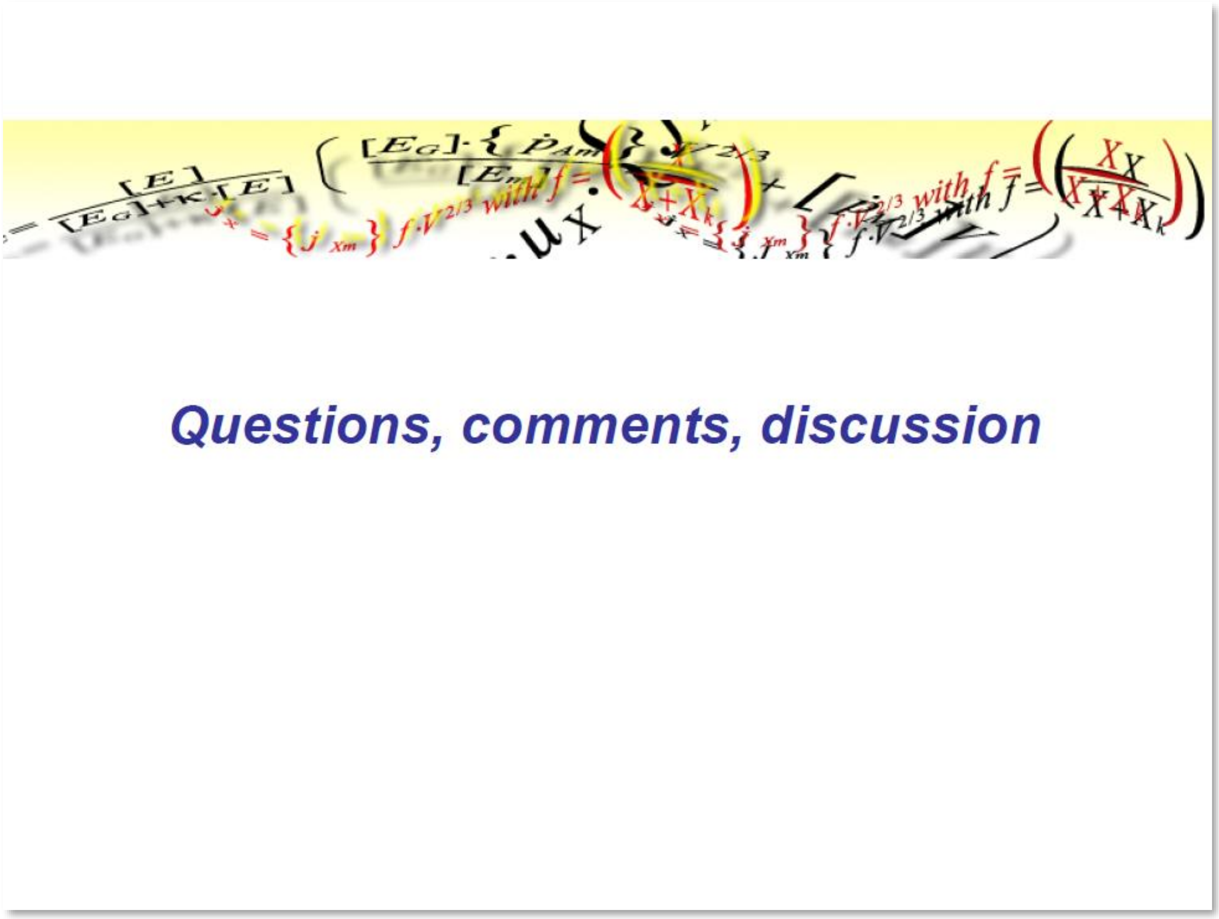
Comparison with measurements

- Verification/validation of model is impossible here
  - only data at end of test; food level was tweaked to match that
  - model predicts very little effect, and data confirm that …

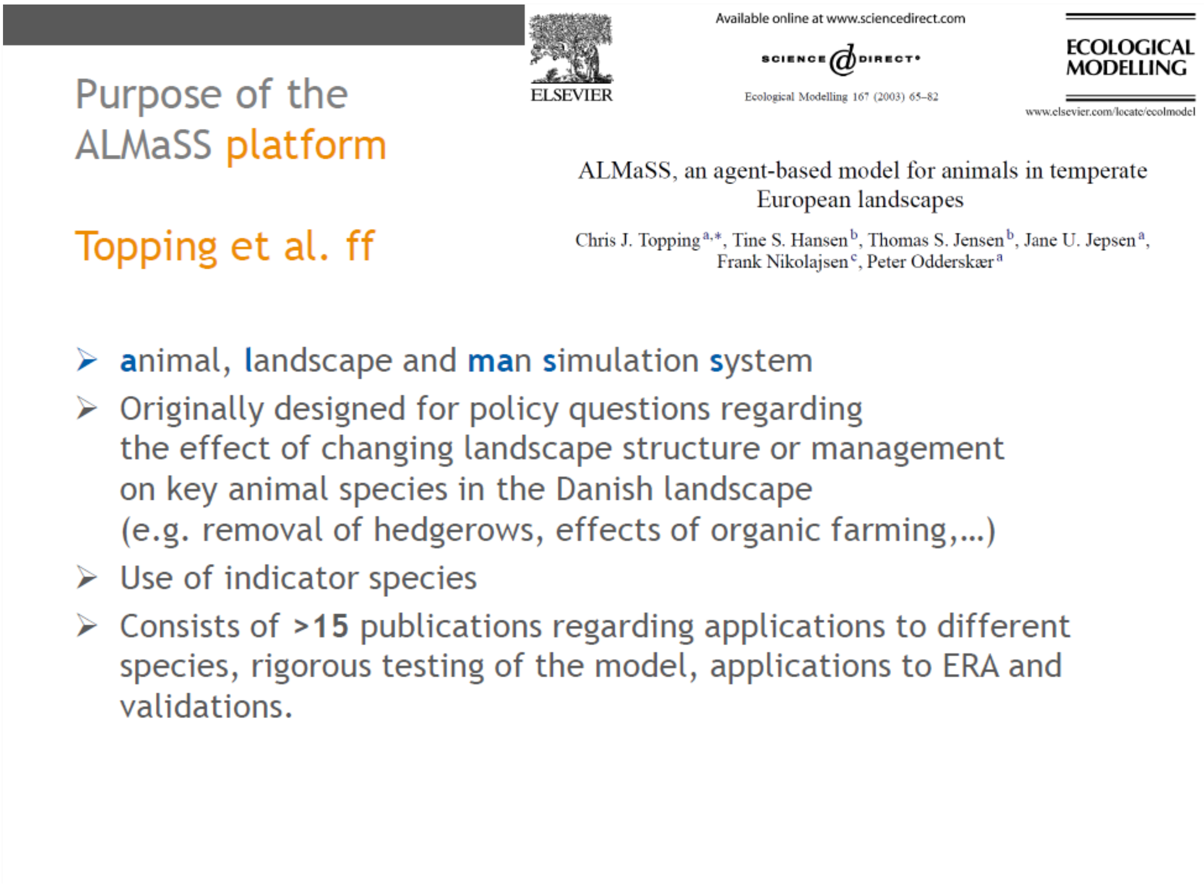


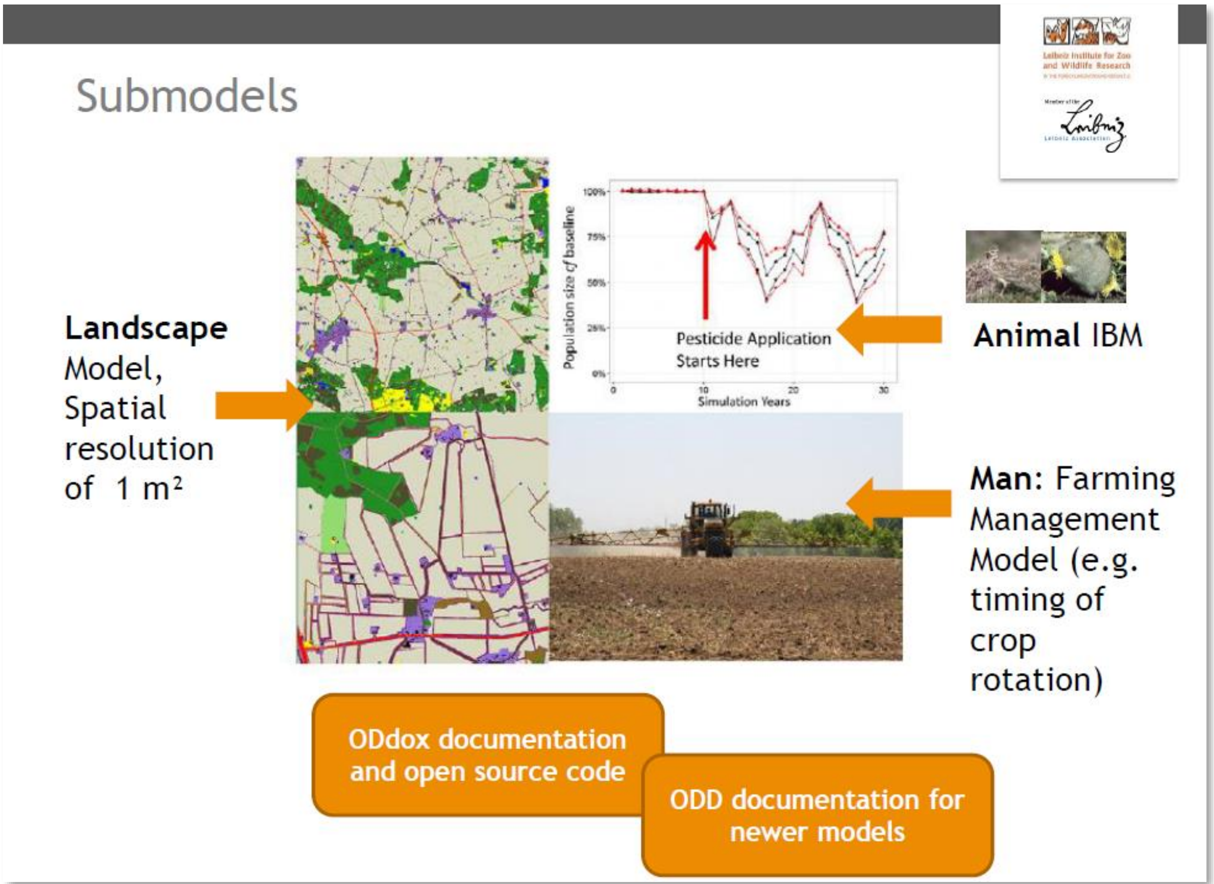
Conclusions on case study

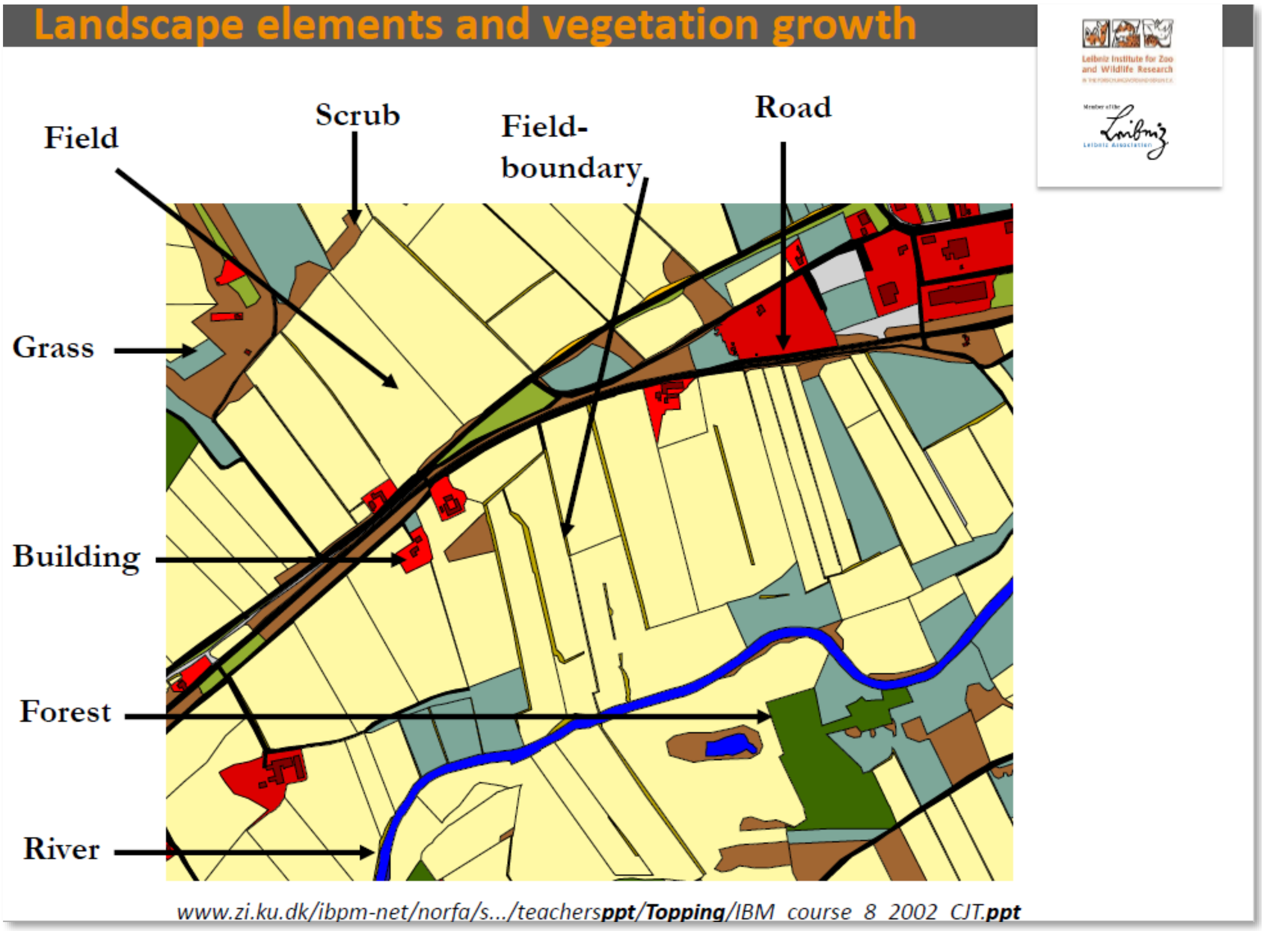
- Author's claims unsupported by analysis:
  - "… model can be used in the context of ecological risk assessment for beta-cyfluthrin"
  - "… we can use the model to make predictions for untested exposure scenarios"
- Most importantly:
  - data from standard ELS insufficient to calibrate/test models
  - lack of case studies as basis for applicability to ELS tox
  - not demonstrated that extrapolation to pulses is possible
- Combination stdDEB/AmP is promising …
  - alternative: simplified DEB models

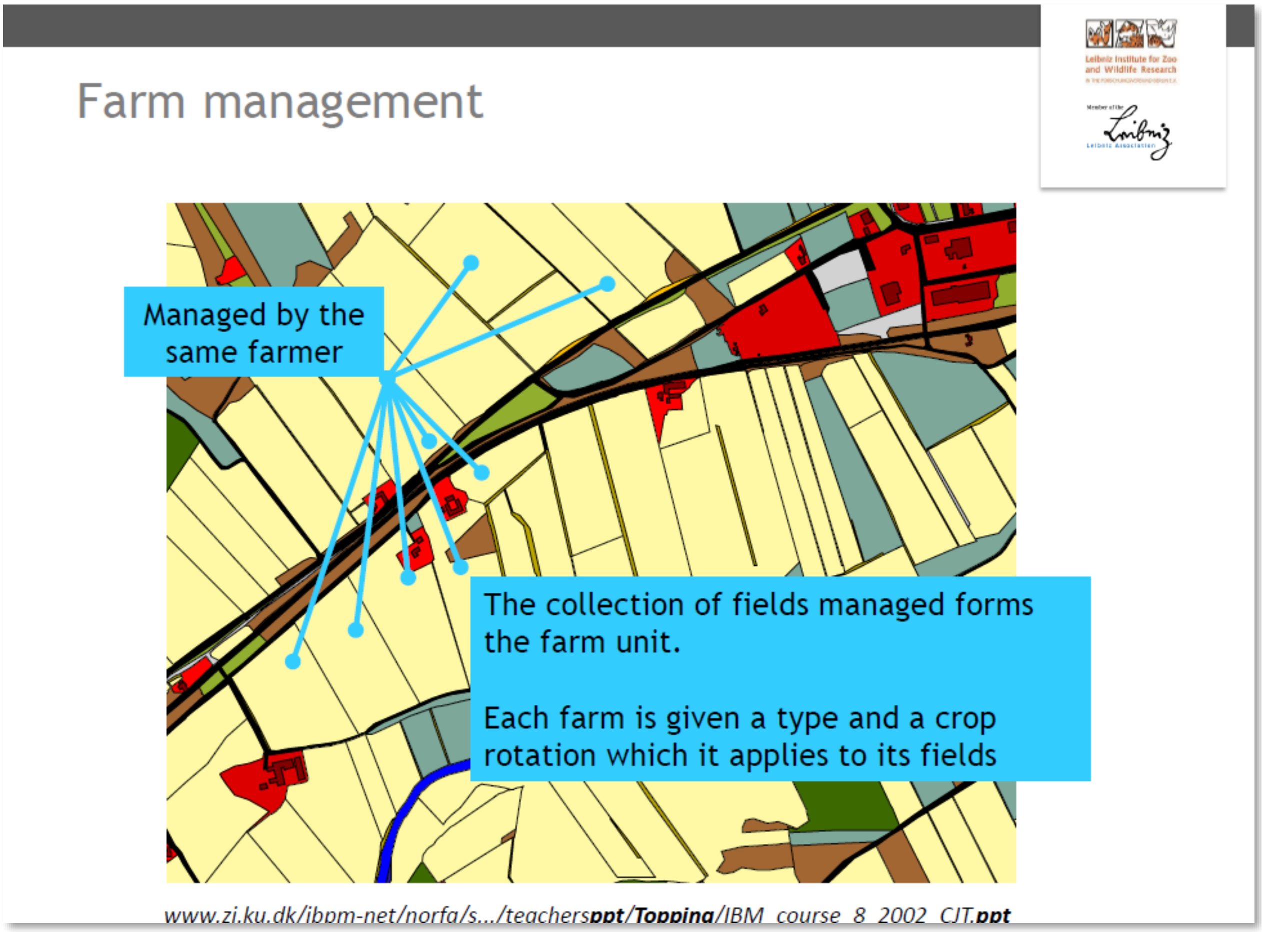

Questions, comments, discussion

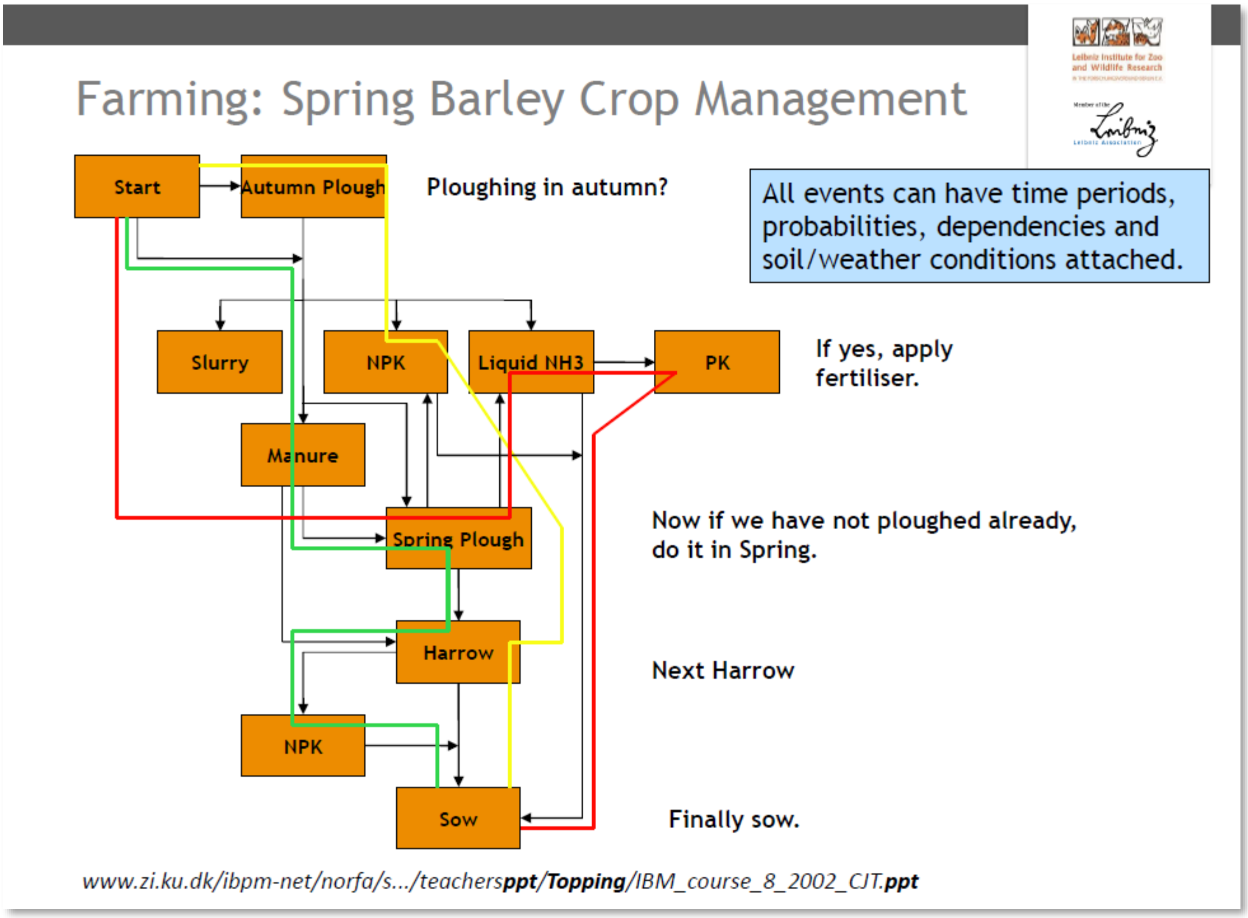
Presention on ALMaSS

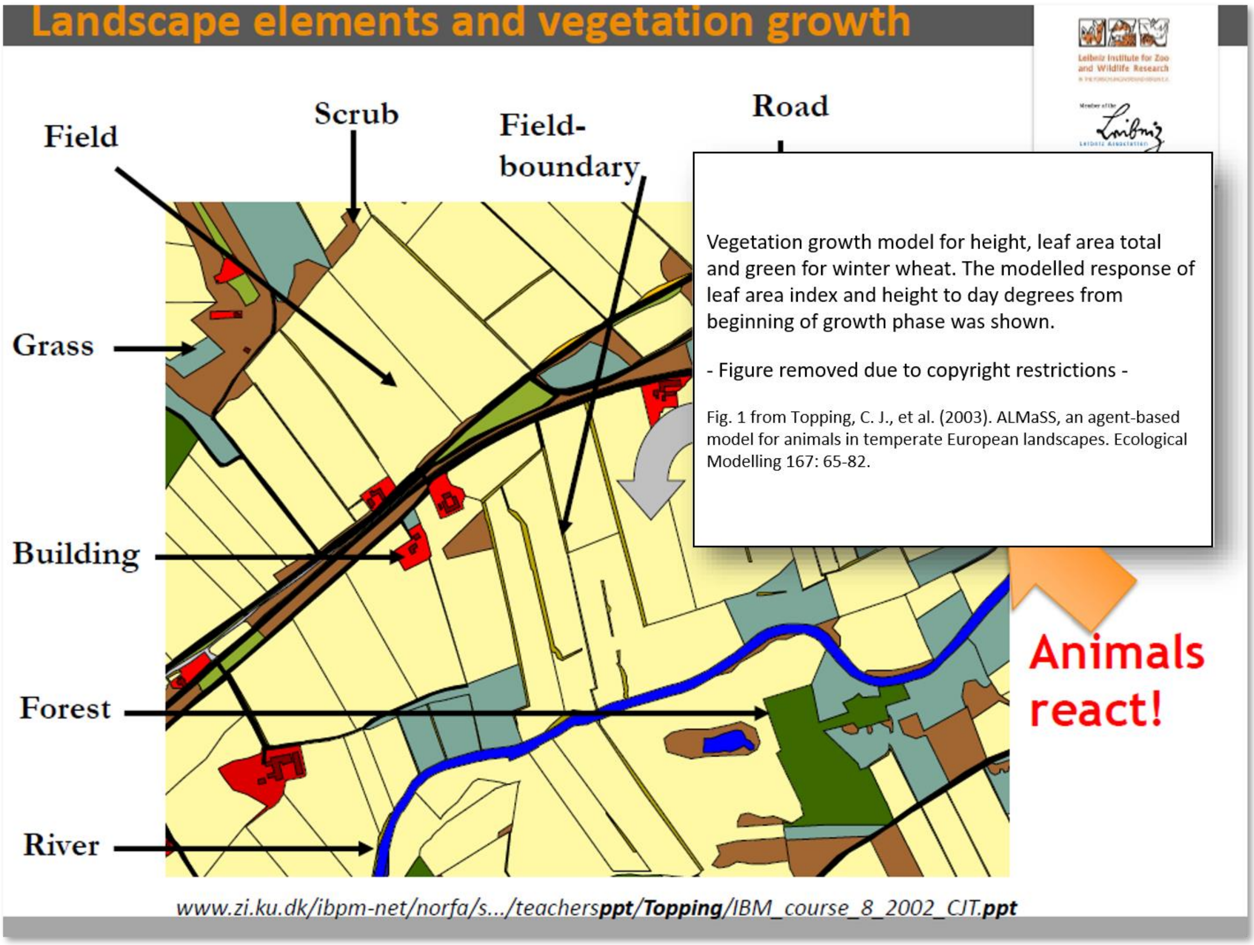





Landscape elements and vegetation growth
Leibniz Institute for Zoo and Wildlife Research
Field
Scrub
Field-boundary
Road
Grass
Building
Forest
River
www.zi.ku.dk/ibpm-net/norfa/s.../teachersppt/Topping/IBM_course_8_2002_CJT.ppt


Farm management
Leibniz Institute for Zoo and Wildlife Research
Managed by the same farmer
The collection of fields managed forms the farm unit.
Each farm is given a type and a crop rotation which it applies to its fields
www.zi.ku.dk/ibpm-net/norfa/s.../teachersppt/Topping/IBM_course_8_2002_CJT.ppt


Farming: Spring Barley Crop Management
Start
Autumn Plough
Ploughing in autumn?
All events can have time periods, probabilities, dependencies and soil/weather conditions attached.
Slurry
NPK
Liquid NH3
PK
If yes, apply fertiliser.
Manure
Spring Plough
Now if we have not ploughed already, do it in Spring.
Harrow
Next Harrow
NPK
Sow
Finally sow.
www.zi.ku.dk/ibpm-net/norfa/s.../teachersppt/Topping/IBM_course_8_2002_CJT.ppt


Landscape elements and vegetation growth
Field
Scrub
Field-boundary
Road
Grass
Building
Forest
River
Vegetation growth model for height, leaf area total and green for winter wheat. The modelled response of leaf area index and height to day degrees from beginning of growth phase was shown.
- Figure removed due to copyright restrictions -
Fig. 1 from Topping, C. J., et al. (2003). ALMaSS, an agent-based model for animals in temperate European landscapes. Ecological Modelling 167: 65-82.
Animals react!
www.zi.ku.dk/ibpm-net/norfa/s.../teachersppt/Topping/IBM_course_8_2002_CJT.ppt

Animals

- Indicator species
- Life-history based on behavioral states and transitions between them
- Decisions as reaction to farming, weather, conspecifics and predation

Open field and marginal habitat birds

Small mammal herbivore, grassland specialist

Large mammal, mosaic specialist

Polyphagous predator 1

Large mammal, woodland mosaic specialist

Polyphagous predator 2

Medium sized mammal

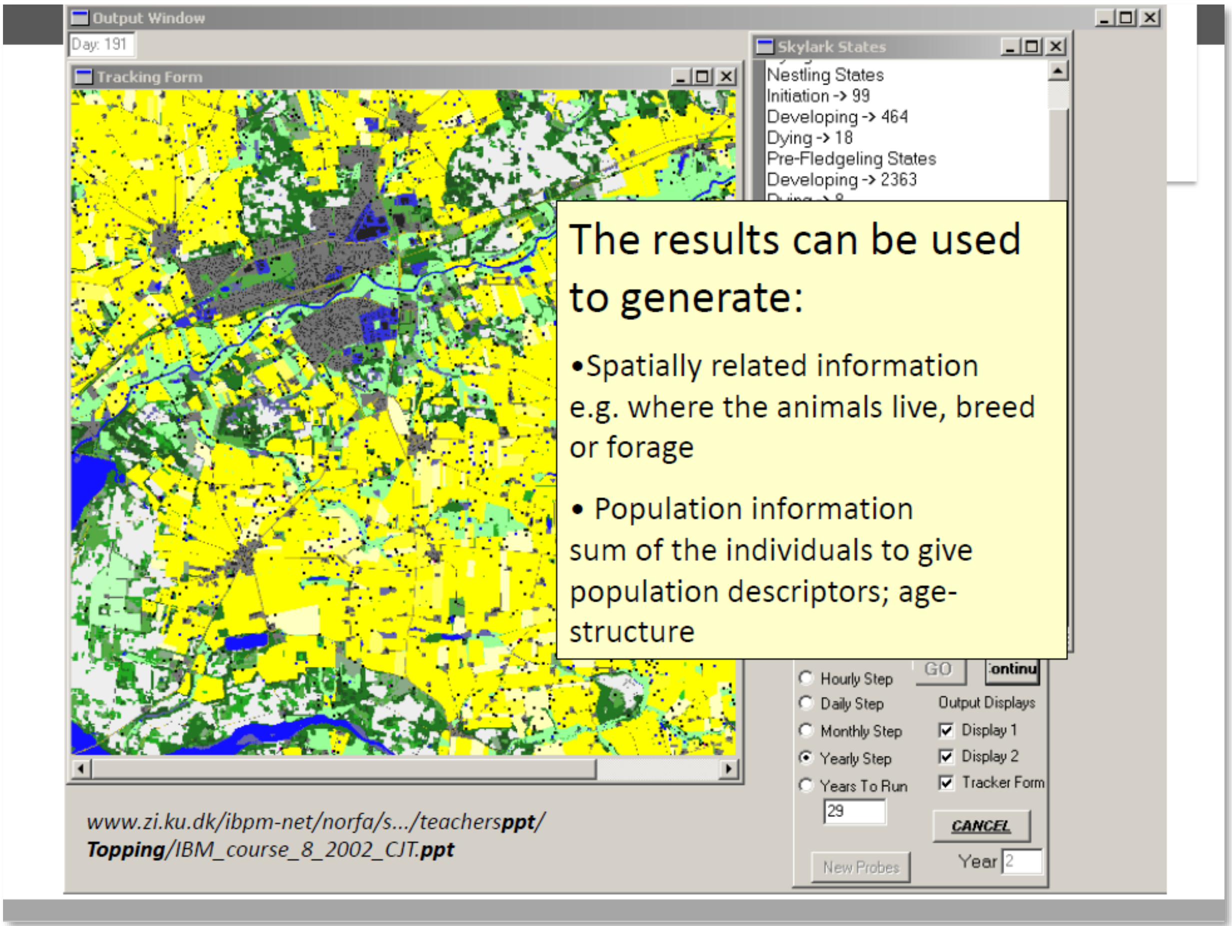
Animal IBMs

- Life-history stages and behavioral states

J.U. Jepsen et al. / Agriculture, Ecosystems and Environment 105 (2005) 581–594 585

Table 1
Species models with behavioural states (transient behaviours may differ for the same object types)

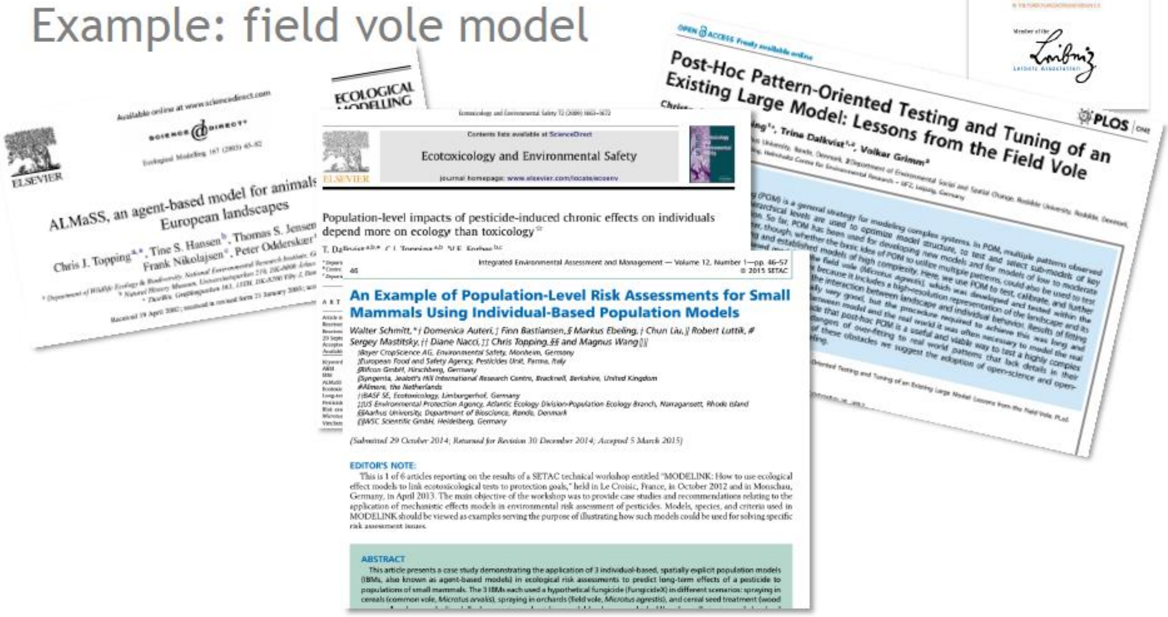
Species model	Object types	Behavioural states
Carabid beetle (*B. lampros*)	Egg (E) Larva (L) Pupa (P) Adult female (♀)	Initiation (E, L, P, ♀), development (E, L, P), hatching (E, L), larval dispersal (L), emerging (P), reproduction (♀), dispersal (♀), aggregating (♀), hibernating (♀), dying (E, L, P, ♀)
Linyphiid spider (*O. fuscus*)	Egg (E) Juvenile (J) Adult female (♀)	Initiation (E, J, ♀), development (E, J), hatching (E), maturation (J), dispersal (J, ♀), assess habitat (J, ♀), assess food (J, ♀), reproduction (♀), dying (E, J, ♀)
Skylark (*A. arvensis*)	Clutch (C) Nestling (N) Prefledgeling (F) Adult male (♂) Adult female (♀)	Initiation (C, N, P, ♂, ♀), developing (C, N, P), hatching (C), leave nest (N), maturation (P), migrating (♂, ♀), flocking (♂, ♀), floating (♂, ♀), find territory (♂, ♀), establish territory (♂), give up territory (♂, ♀), build-up resources (♀), start new brood (♀), attract mate (♂, ♀), mate guarding (♂), make nest (♀), lay eggs (♀), incubation (♀), parental care (♂, ♀), dying (C, N, P, ♂, ♀)
Field vole (*M. agrestis*)	Adult male (♂) Adult female (♀)	Initiation (♂, ♀), evaluate and explore (♂, ♀), assess habitat (♂, ♀), dispersal (♂, ♀), maturation (♂, ♀), lactation (♂, ♀), giving birth (♀), mating (♀), infanticide (♂), dying (♂, ♀)
Roe deer (*C. capreolus*)	Juvenile (J) Adult male (♂) Adult female (♀)	Initiation (J, ♂, ♀), maturation (J), assess habitat (♂, ♀), feeding (♂, ♀), ruminating (♂, ♀), dispersal (♂, ♀), establish range (♂, ♀), establish territory (♂), social ranking (♂), mating (♂, ♀), give birth (♀), winter grouping (♀), dying (J, ♂, ♀)



Application to ERA – Pesticide application

Relevant application examples	Taxa	Chemicals	Comments
Topping & Odderskjaer (2004).MODELING THE INFLUENCE OF TEMPORAL AND SPATIAL FACTORS ..... Environmental Toxicology and Chemistry, Vol. 23, No. 2, pp. 509–520, 2004	Skylark ✓	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
Topping et al. 2005. Risk Assessment of UK Skylark ..... Ecotoxicology, 14, 925–936, 2005	Skylark	pesticide	a comparison of a non-spatial IBM with ALMaSS handling and outcome; shows advantage of ALMaSS flexibility
Dalkvist et al. 2009. Population-level impacts of pesticide-induced chronic .... Ecotoxicology and Environmental Safety 72 (2009) 1663–1672	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. RECOVERY BASED ON PLOT EXPERIMENTS... Environmental Toxicology and Chemistry, Vol. 33, No. 7, pp. 1499–1507, 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.
Topping et al. 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	brown hare ✓	Insecticide (fictious endocrine disruptor)	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.
Topping et al. 2015. Towards a landscape scale management of pesticides: ERA using changes in modelled occupancy and abundance .... Science of the Total Env. 537 (2015) 159–169	carabid beetle (Bembidion lampros), a linyphiid spider (Oedothorax fuscus)	insecticide	Pesticide stressors are simulated as changing spatial and temporal concentrations, based on spraying regimes and environmental fate of the active substances.

Example: field vole model



Topping et al. 2003
Ecological Modelling 167

Dalqvist et al. 2009
Ecology and Env Safety 72
Schmitt et al. 2015
Int Env Ass Manage 12

Topping et al. 2012
PlosOne 7

Description | **Application in science** | **Evaluation basic model**

Example: field vole

Modelink workshop

46 Integrated Environmental Assessment and Management — Volume 12, Number 1—pp. 46–57
© 2015 SETAC

An Example of Population-Level Risk Assessments for Small Mammals Using Individual-Based Population Models

Walter Schmitt,† Domenica Auteri,‡ Finn Bastiansen,§ Markus Ebeling,† Chun Liu,|| Robert Luttik,# Sergey Mastitsky,†† Diane Nacci,‡‡ Chris Topping,§§ and Magnus Wang||||*

Application of FungicideX → deterministic long-term first tier risk assessment resulting in high risk to small mammals

Criteria:

- A population density 5% below control considered negligible
- No long-term decline of population density during consecutive use of product

Study species

Parameterization for *Microtus agrestis* („field vole“)

Microtus agrestis

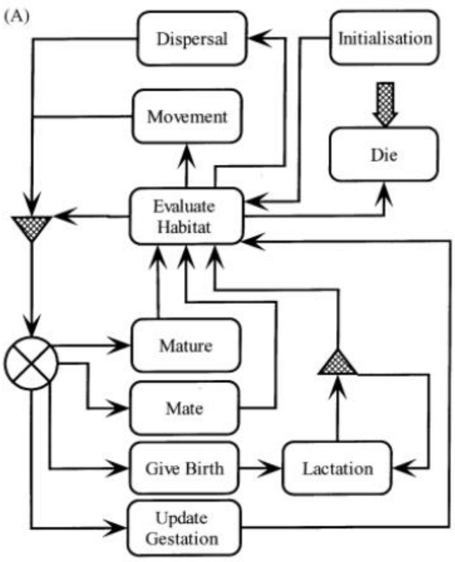
Characteristics of *M. agrestis*

- Common species
- Prefers tends to prefer wet areas (marshes, bogs and river banks)
- Feeds on grasses, herbs, root tubers, moss and other vegetation
- important part of the diet of owls/raptors/ weasels/foxes → ES
- breeds throughout the year,
  but the breeding season peaks in spring and summer
- Lives in shallow burrows
- Nest sharing of females, but territorial
- Males disperse in search of territories/ females
- Typical r strategist (mass development)
- Short generation time (21 d), many pups

Surrogate species for small herbivorous mammals (EFSA birds & mammals guidance, 2009)

The vole model (Topping et al. 2003)

Field vole (*M. agrestis*)	Adult male (♂) Adult female (♀)	Initiation (♂, ♀), evaluate and explore (♂, ♀), assess habitat (♂, ♀), dispersal (♂, ♀), maturation (♂, ♀), lactation (♂, ♀), giving birth (♀), mating (♀), infanticide (♂), dying (♂, ♀)



Female cycle

Territory establishment and movement/dispersal are based on habitat quality (aggregated values from land cover types)

If habitat quality below a threshold, body condition decreases, otherwise increases (simple energy budget proxy)

Giving birth is dependent on habitat condition

Population dynamics emerge from territory dynamics;
density regulation via space (territory)

POM - Model evaluation

OPEN ACCESS Freely available online PLOS ONE

Post-Hoc Pattern-Oriented Testing and Tuning of an Existing Large Model: Lessons from the Field Vole

Christopher J. Topping[1*], Trine Dalkvist[1,2], Volker Grimm[3]

Good fit of model results to data

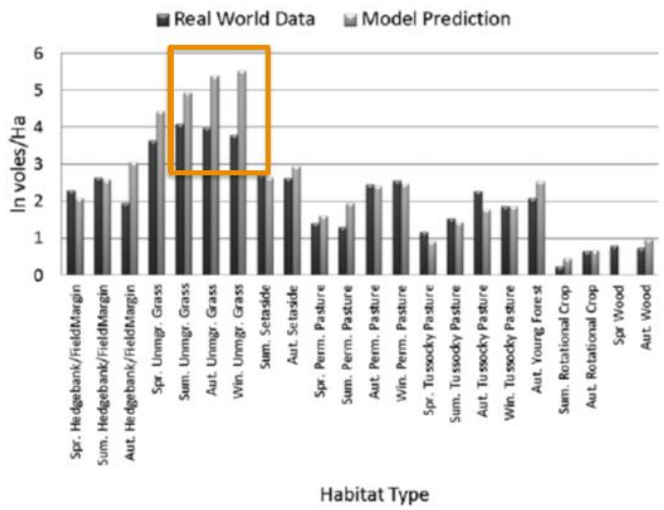


Figure 6. Real world means and model means for total vole density for a range of Danish habitats and sampling periods. X-axis abbreviations: Spring (Spr.), Summer (Sum.), Autumn (Aut.), Winter (Win.), Unmanaged (Umgr.) Permanent (Perm.).
doi:10.1371/journal.pone.0045872.g006

The case: Fungicide X

Specific individual-level endpoints like body-weight of parents and offspring, litter size etc were translated into population-level effects.

Dose-response relationships on reproductive performance (based on toxicological results from 15 different pesticides)

Table 1. Toxicological profile of FungicideX used in the case study risk assessments

		Parental F0 females			F1 pups					
	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)	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)	20	100.4	100.1	100.7	101.2	100.0	100.0	98.3	0.3	-0.3
LOAEL (mid dose)	50	97.6	96.9	97.7	92.9	96.2	96.6	93.2	1.0	0.7
High dose	150	91.7	90.7	89.7	71.8	71.1	86.9	67.8	2.4	2.2

F0, F1: Denotation of generations in multi generation studies; NOAEL: No observable adverse effect level; LOAEL: Lowest observable adverse effect level.

Schmitt et al. 2015

The case: Fungicide X

- Effects on reproductive performance simulated
- Simple dose-response function for 3 endpoints finally modelled
- Scaling linear function between 0 effect (< NOAEL) and 100% effect (high dose; resulting in 50% population reduction)

	Dose level [mg/kg bw/d]
NOAEL (low dose)	20
LOAEL (mid dose)	50
High dose	150

Table 4. Coefficients *a* and *b* for the linear dose–response curves of the 3 types of sublethal effects when 20 mg a.i./kg bw < effective-dose ≤ 150 mg a.i./kg bw

	Effect = *a* × Effective-dose + *b*		
Coefficient	**Litter size (%Ctrl)**	**Pup survival (%Ctrl)**	**Delay of 1st reproduction (days)**
a	−0.0022	−0.0023	0.0156
b	1.0494	1.0593	0.0885

Schmitt et al. 2015

The case: Fungicide X

- Effects on reproductive performance simulated
- Simple dose-response function for 3 endpoints finally modelled

- Models calculated individual daily exposures based on the substance residues varying spatially in the landscape
- no TK: rapid elimination of FungicideX from animal's body assumed; no bioaccumulation
- TD simplified: effects are induced immediately when the critical dose is achieved, and recovery of health occurs when the dose drops below the threshold

NOAEL (low	
LOAEL (mid	
High dose	150

Schmitt et al. 2015

The case: Spraying of orchards

Emergence from model processes: Spatial choice of feeding and the time spent feeding in the treated habitat considered

2.5% treated orchards as main habitat for voles in landscape, i.e. most voles inhabited treated orchards
Non-orchard areas not treated

Simulation of 20 pre-treatment, 10 years treatment + 20 post-treatment
ApplicationRate 1,5,10x

Three 10 x 10 km landscapes used for scenario simulation of field vole populations with ALMaSS for a case study in the SETAC workshop MODELINK.

- Figure removed due to copyright restrictions -

Fig. 2 from Schmitt, W., et al. (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.

Table 2. Hypothetical agricultural use pattern (GAP) used in the case study risk assessments

Crop	Timing of application defined by growth stage*	Number of applications per season	Maximum application rate
Pome fruits (spraying orchard application)	BBCH 51–55	1	0.85 kg a.s./ha

Schmitt et al. 2015

The case: Results compared to control

Predicted effects on field vole populations for the three landscape scenarios shown on the previous slide.

Population size over time is shown for the treated orchards only and for the full 10 x 10 km modelled landscape. 7 different exposure scenario were run. Devation from control population size is clearly visible only for the toxic standard exposure scenario (50 acute mortality of exposed individuals).

- Figure removed due to copyright restrictions -

Fig. 6 from Schmitt, W., et al. (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.

Only the toxic standard treatment resulted in any effect!

Slow recovery of empty orchards due to marginal dispersal of species yielded effects > 10 years

Low risk concluded

However, community effects on predator population not accounted for.

Schmitt et al. 2015

ALMaSS – Strength and limitation

- **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; well documented and parts rigorously tested and validated (vole, skylark, hare)

- **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. intersepcific 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** So far, no chronic effects on survival and reproduction after pulse exposure assumed (depending on model).

- **Conclusions** The only model of its kind that deals so flexibly with landscapes, farm management and agent-based models. Great for scenario comparison. ALMaSS studies emphasise the need for greater focus on animal ecology in risk assessments.

Jeremias Becker, Tjalling Jager, Stephanie Kramer-Schadt, Mathias Franz, Matthias Liess, Magali Solé, Sabine Duquesne, Silvia Pieper, Steffen Matezki, Tobias Frische, Jörn Wogram

Thank you!

Umwelt Bundesamt

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

DEBtox Research

Leibniz-Institut für Zoo- und Wildtierforschung

Presentation on the IBM *Chaoborus* Population Model

Chaoborus IBM Population Model

gaiac Research Institute for Ecosystem Analysis and Assessment

Tido Strauss et al. (2016)

Characteristics

3

Characteristics

- Individual based model
- Interactions between different life stages (cannibalism)
- Metapopulations
- Connection to LC50 model or GUTS for individual-level effects
- Time steps 1d
- Few rules and parameters

Objectives

- „Predicting population dynamics of *Chaoborus chrystallinus* at different food levels and fluctuatingtemperature and light conditions in outdoor ponds based on individual life-cycles" (Strauß et al. 2016).
- Extrapolating ecotoxicological effects from individual to population level.
- Demonstrating "that substances or products ... do not have ... any unacceptable effects on the environment" (EC Regulation No. 1107/2009).

Study species 4

Chaoborus chrystallinus („phantom midge“, „glassworm“)

Life cycle of *Chaoborus crystallinus*

- Figure removed due to copyright restrictions -

Fig. 1 from Strauss, T., et al. (2016). The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

Strauss et al. (2016)

Characteristics

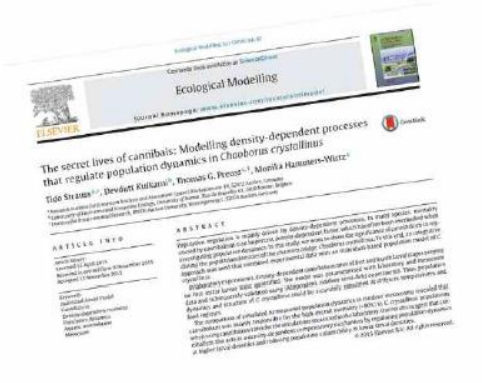
- In fish-free ponds
- Highly sensitive to insecticides
- Tolerant to organic pollution
- 2-3 generations / year (temperate zone)
- Larval hibernation
- Opportunistic predators

  L1/L2: Rotifera etc.
  L3/L4: Cladocera etc.
- Cannibalism potentially common
- Adults may colonize adjacent waters

Medium vulnerability

Not protective for univoltine species

Publications on the model 5


Ecological Modelling

The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus

Strauss T, Kulkarni D, Preuss TG, Hammers-Wirtz M.

The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in *Chaoborus crystallinus*.

Ecological Modelling (2016), *321*, 84-97.

https://doi.org/10.1016/j.ecolmodel.2015.11.004

Description


Population-Level Effects and Recovery of Aquatic Invertebrates after Multiple Applications of an Insecticide

Dohmen GP, Preuss TG, Hamer M, Galic N, Strauss T, Van den Brink PJ, De Laender F, Bopp S.

Population-level effects and recovery of aquatic invertebrates after multiple applications of an insecticide.

Integrated Environmental Assessment and Management (2016), *12* (1), 67-81.

https://doi.org/10.1002/ieam.1676

Evaluation

Model documentation

6


Documentation of the IBM Chaoborus Model V4.1.2 – Tido Strauss, Research Institute gaiac 1

Description of the individual-based model
"IBM *Chaoborus* population model"
for the aquatic phantom midge *Chaoborus crystallinus*
Documentation of the basic population model
including a metapopulation approach for use in environmental risk assessment

Tido Strauss

Research Institute for Ecosystem Analysis and Assessment (gaiac)
Kackertstrasse 10, 52072 Aachen, Germany
strauss@gaiac.rwth-aachen.de

January 2017

This model description applies to the *Chaoborus crystallinus* population model (version 4.1.2) and follows the ODD (Objects, Design concepts, and Details) protocol (Grimm et al. 2006, 2010).

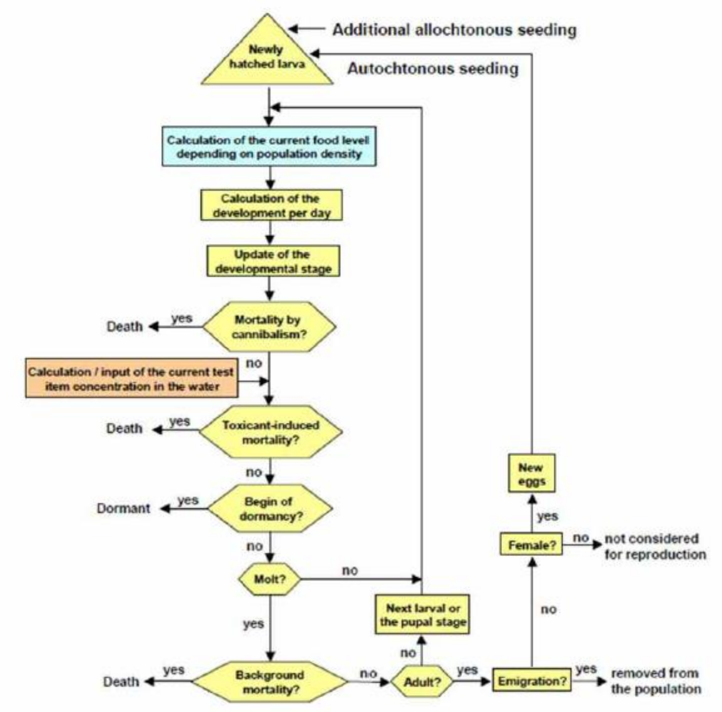
ODD standard

Grimm et al. (2006, 2010)

Strauss T (2017)

Model mechanisms

7



Strauss & Norman (2017)

Model parameters

8

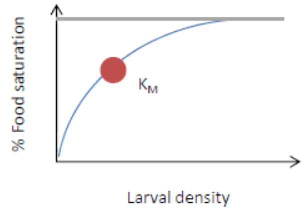
Life history parameters

From random distribution at birth

Sex
Individual development rate
Individual factor for background mortality (renewed after each moulting)
Susceptibility to cannibalism (daily renewal)
Susceptibility to dormancy

Fixed

Clutch size: 160
Clutch per female: 0.4
Background mortality: 20%
K_M for density-dependent food availability: 0.8



Environmental parameters

Volume (size + depth) of water body
Food amount (constant)
Seasonal pattern of temperature, photoperiod
Immigration + emigration rate
PPP application time, concentration, dissipation rate
Coupling to individual-level effects

Main processes

9

Larval development time depends on

Individual development rate, sex, temperature, larval density (L1/L2 or L3/L4).
No development during dormancy or at < 6°C

Food saturation [%] vs. prey/L (rotifers and nauplii, or cladocerans)

- Figure removed due to copyright restrictions -

Fig. 4 from Strauss, T., et al. (2016). The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

Strauß et al. 2016

Mortality from cannibalism depends on

Temperature (L3/L4), density of predator (L3/L4) and prey (L1/L2)

Mortality [%/d] and intake [ind./L] vs. density of first instar larvae [ind./L]

- Figure removed due to copyright restrictions -

Fig. 3 from Strauss, T., et al. (2016). The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

Strauß et al. 2016

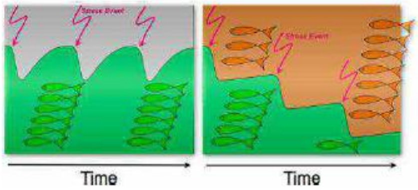
Developmental success depends on

Larval density + temperature (indirectly)

Main processes

10

Concerns



Liess et al. (2013)

- **No species interactions**

  ⇨ Predictions only for isolated populations (no predators + competitors)
  No indirect food-web effects

- **No additional environmental stressors**

  ⇨ Drought, other pollutants etc. may increase development time + background mortality
  Predictions rather for mesocosm than field conditions

- **Cannibalism modeled rather like competition**

  ⇨ No gain for predator
  No indirect effects on predators via prey reduction
  All but L2 affect mortality of L1

Sensitivity analysis

11

Sensitivity for five endpoints (mean larval abundance, laraval abundance at and, emerged adults, eggs, dead larvae) to eleven parameters or processes, respectively (development rate, temp. coeff. develop., background mortality, cannibalism rate L1, cannibalism rate L3/L4, egg number per clutch, clutches per female, half sat. const. food, initial density, food sat. L1/L2, food sat. L3/L4)

- Figure removed due to copyright restrictions -

Fig. B1 from Strauss, T., et al. (2016). The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

Strauß et al. (2016).

Sensitivity coefficient

> 1: Change of endpoint > change of parameter

< 1: Change of endpoint < change of parameter

< 0: Negative correlation of parameter and endpoint

Validation of basic population model

12

Measured and simulated population size of a) L1 larvae and eggs, b) L2 larvae, c) L3 larvae and d) L4 larvae vs. day of the year.

a) The mean simulated L1 and egg populaton size increased from (close to) zero at day of the 160 to 10,000 (range: 5,000 – 17,000) at day 190 before decreasing again to < 2,000 by day 220. The observed number of larvae and eggs fall within the same range in the same season.

b) The mean simulated number of L2 larvae increased from < 500 at day of the year 160 to 5,000 200 (range: 2,500 – 7,000) at day before decreasing again to < 500 by day 230. The observed number of L2 larvae was in the same range but reached their peak later (around day 240). Around day 200, < 500 L2 larvae have been obsereved in most cases (but one data point shows 6,000 L2 larvae).

c) The mean simulated number of L3 larvae increased from ca. 1,000 at day of the year 160 to 6,000 (range: 4,000 – 8,000) around day 230 before decreasing again to < 2,000 by day 260. The observed L3 population sizes was in the same range, but started with typically higher numbers (around 2,000 individuals) at day 160 and reached their peak not before day 240.

d) The mean simulated number of L4 larvae started with ca. 1,000 at day of the year 160 and increased between day of the 210 to 260 to 5,000 (range: 2,000 – 7,000) individuals. The observed L4 population sizes generally matched predictions but showed also some higher values outside of the predicted range throughout the observed period (up to 2,000 individuals at the beginning, up to 8,000 individuals at the end).

- Figure removed due to copyright restrictions -

Fig. 8 from Strauss, T., et al. (2016). The secret lives of cannibals: Modelling density-dependent processes that regulate population dynamics in Chaoborus crystallinus. Ecological Modelling 321: 84-97.

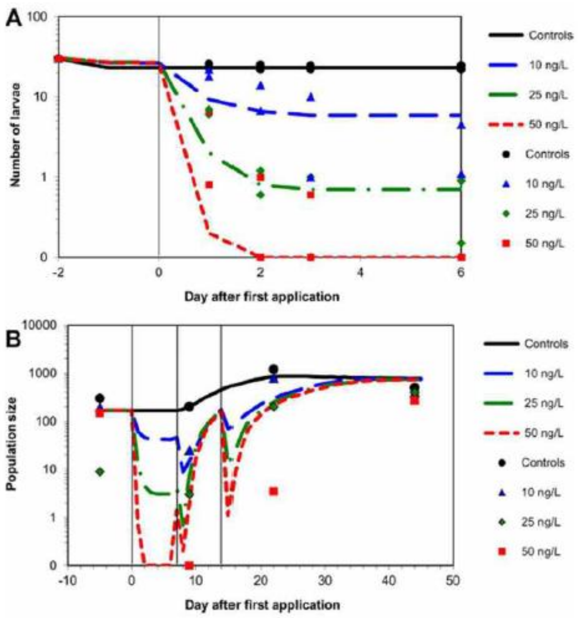
Strauß et al. 2016

Validation of predicted recovery

13

MODELINK workshop

- Application of „modelmethrin" on
  - Chaoborus IBM
  - MASTEP (*Gammarus pulex*)
  - IDamP (*Daphnia magna*)
- GUTS for individual-level effect
- All models provide similar risk



Demonstration that the model is not overly conservative.

No prediction of long-term effects when they have not been observed.

But is the model conservative enough?

Prediction of long-term effects when they have been observed?

Dohmen et al. 2016

Case study

14

Strauss T, Norman S (2017)
Modelling Chaoborus crystallinus populations to simulate effects and recovery under beta-cyfluthrin exposure
gaiac - Research Institute for Ecosystem Analysis and Assessment
ADAMA Report No. R-37809

Background

- EU Annex I Renewal of beta-cyfluthrin (spraying on cerals + potatoes)
- PEC_{SW} in D2 ditch = 3 ng/L and in D4 pond = 1 ng/L (FOCUS step 4 with 20 m buffer strip)
- Mesocosm study (Jenkins, 2014)

  Chaoborus sp. (**LOEC = 1.6 ng/L**)
  Crangonyx pseudogracilis, amphipoda (NOEC = 5 ng/L)
  Less sensitive: plecoptera, trichoptera, megaloptera
- ⇨ Without *Chaoborus*: ETO-RAC = 2.5 ng/L (AF = 2; achievable with drift reducing nozzle)
- Modelling with $PEC_{assumed}$ = 2.5 ng/L (ditch) or 0.75 ng/L (pond)

Idea

Deriving the ETO-RAC from *Crangonyx* instead of *Chaoborus* by demonstrating that *Chaoborus* recovers in < 8 weeks.

Case study

15

Environmental scenario

Exposure

- 2 spray applications (7.5 g/ha) with 14 d interval at May 15 + 29 (benchmark: June 6)
- 2.5 ng/L (D2 ditch) or 0.75 ng (D4 pond)

Starting conditions

- January 1st; 1 y simulation in 1 d steps
- 100 % L4 larvae
- 1,000 – 3,000 individuals (1 - 3 larvae/L in a 1 m^3 pond)

Other conditions

- Ditch: 50 % migration between exposed and control
- Ditch: Reference without migration
- Pond: 50 % emigration of adult females, no immigration (worst case)
- Temperature and food based on mesocosm studies

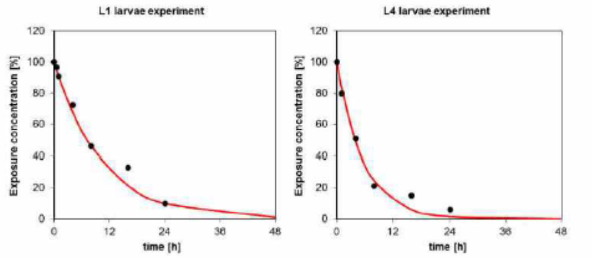
Case study

16

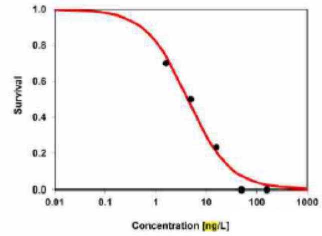
Individual-level effects

LC50 approach (DT_{50} < 1 d)

- Only acute mortality from 2 tests
- L1 without sediment (72 h)
  DT_{50} = 7.2 h
  All dead at 1.6 ng/L
- L4 with sediment (96 h)
  DT_{50} = 3.9 h
  2.5ng/L -> 36 % mortality
  0.75 ng/L -> 14 % mortality (extrapolated)
  Recovery of "ecologically dead" possible



Dissipation with (left) and without (right) sediment (Strauss & Norman 2017).



Dose response curve L4 96 h (Strauss & Norman 2017).

Implementation

- DT_{50} = 4 h
- 2.5 ng/L pulse: 100 % mortality L1/L2, 36 % mortality L3/L4
- 0.75 ng/L pulse: 100 % mortality L1/L2, 14 % mortality L3/L4
- Unrealistic worst case (?)

Case study

17

Modeling results

D2 ditch: 50 % migration between exposed and control

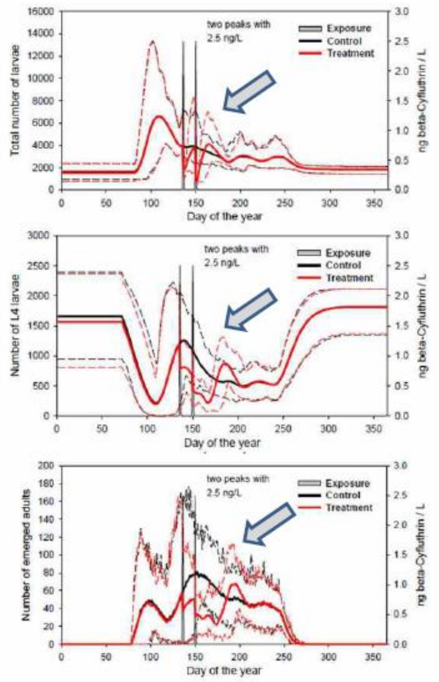
Recovery time after 1st pulse

All larvae: 24 d

L4: 41 d

Adults: 49 d

Recovery just within < 8 weeks.



Mean, min and max from 100 simulations (Strauß & Norman 2017).

Case study 18

Modeling results

D2 ditch: Reference without migration

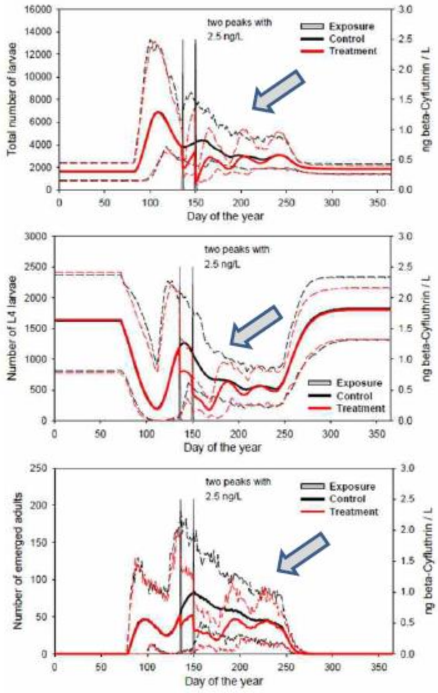
Recovery time after 1st pulse

All larvae: 70 d

L4: 50 d

Adults: 90 d

Recovery > 8 weeks due to missing recolonization.



Mean, min and max from 100 simulations (Strauß & Norman 2017).

Case study 19

Modeling results

D4 pond: 50 % emigration

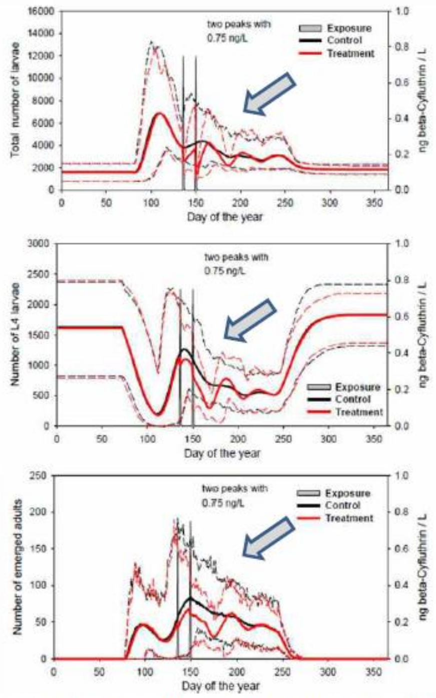
Recovery time after 1st pulse

All larvae: 28 d

L4: 44 d

Adults: 54 d

Slowed down recovery due to missing recolonization, even though exposure was lower than in ditch.



Mean, min and max from 100 simulations (Strauß & Norman 2017).

Case study

20

Conclusions

Uncertainties leading to overestimation of risk

- 100 % mortality L1/L2 (without sediment too long exposure)
- No immigration in pond

Uncertainties leading to underestimation of risk

- No species interactions
- Few additional environmental stressors
- No sublethal effects
- High L1/L2 mortality does not result in starvation of L3/L4
- Uncertainty in L3/L4 mortality
- Ditch: Control was indirectly affected by pesticides
- Ditch: Immigration rate possibly too high
- Potentially reduced recovery following applications later in the season

Potential of unbalanced risk assessment!

Summary

21

Strengths

- Model mechanisms transparent, not too complex
- Validation of population dynamics, attempt to validate predictions on relevant recovery

Concerns on model development

- Cannibalism implemented as -/0 relationship
- No species interactions
- Parameterization and validation potentially not realistic worst case in field

Main concerns on model application

- *C. crystallinus* not representative for univoltine species
- No sublethal / delayed effects
- Control in ditch scenario indirectly affected by PPP
- No sufficient validation before model application

Jeremias Becker, Tjalling Jager, Stephanie Kramer-Schadt, Mathias Franz, Matthias Liess, Magali Solé, Sabine Duquesne, Silvia Pieper, Steffen Matezki, Tobias Frische, Jörn Wogram

Thank you!

Umwelt Bundesamt

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

DEBtox Research

Leibniz-Institut für Zoo- und Wildtierforschung

Leibniz-Gemeinschaft

Presentation on eVole

eVole

RIFCON GmbH

Wang & Grimm (2010), Wang (2013)

Characteristics

3

Objectives

- Proximate purpose: capture the relation between home range dynamics and population dynamics (Wang & Grimm 2007)
- Ultimate purpose: predicting effects of changes in agricultural practice and pesticide risk assessment (Wang & Grimm 2007)
- Extrapolating ecotoxicological effects from individual to population level.
- Demonstrating “that substances or products produced or placed on the market do not have ... any unacceptable effects on the environment” (EC Regulation No. 1107/2009).

Characteristics

- Individual based model
- Spatially explicit home ranges
- Connection to dose-response model for individual-level effects (lethal, sublethal)
- More rules and parameters, simpler equations (population dynamics as emerging property)

Study species 4

Parameterization for *Microtus arvalis* („common vole“)
and initially *Sorex araneus* („common shrew“)

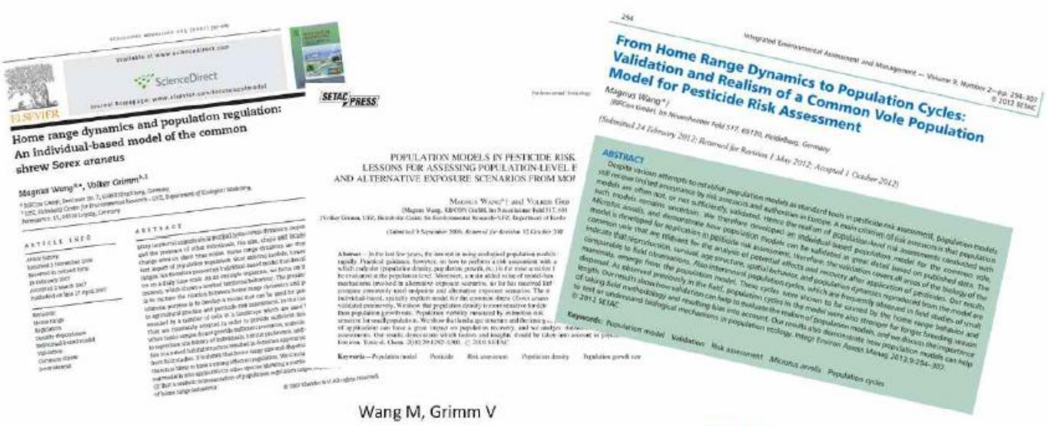
Characteristics of *M. arvalis*

- Prefers open landscape (fields, meadows, pastures)
- Feeds on crops, grass, herbs (day and night)
- Lives in colonies in connected holes with runways
- Nest sharing of females, but territorial
- Males disperse in search of females
- Typical r strategist (cyclic mass development)
- Short generation time (33 d), many pups
- Important prey for raptors, owls and carnivores
- Pest species in agricultural fields

Relatively low vulnerability

Generic focal species for small herbivorous mammals (EFSA birds & mammals guidance, 2009)

Publications on the model 5



Wang M, Grimm V

Home range dynamics and population regulation: An individual-based model of the common shrew Sorex ayaneus

Ecological Modelling (2007), *205* (3-4), 397-409

https://doi.org/10.1016/j.ecolmodel.2007.03.003

Description

Wang M, Grimm V

Population models in pesticide risk assessment: Lessons for assessing population-level effects, recovery, and alternative exposure scenarios from modelling a small mammal

Environmental Toxicology and Chemistry (2010), *29* (6), 1292-1300

https://doi.org/10.1002/etc.151

Application in science

Wang M.

From home range dynamics to population cycles: Validation and realism of a common vole population model for pesticide risk assessment

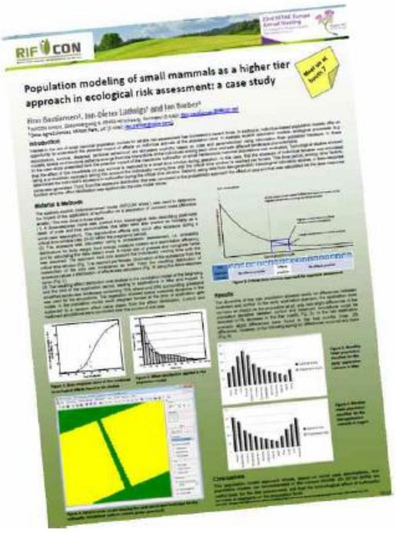
Integrated Environmental Assessment and Management (2013), *9* (2), 294-307

https://doi.org/10.1002/ieam.1377

Evaluation basic model

Publications on the model

6



Bastiansen F, Ludwigs J-D, Barber I

Population modeling of small mammals as a higher tier approach in ecological risk assessment: a case study

Poster at *SETAC* conference (2013)

Application


An Example of Population-Level Risk Assessments for Small Mammals Using Individual-Based Population Models

EDITOR'S NOTE:

ABSTRACT

Schmitt W, Auteri D, Bastiansen F, Ebeling M, Liu C, Luttik R, Mastitsky S, Nacci D, Topping C, Wang M

An example of population-level risk assessments for small mammals using individual-based population models

Integrated Environmental Assessment and Management (2015), *12* (1), 46-57

https://doi.org/10.1002/ieam.1640

Evaluation of recovery predictions

Model documentation

7

TRACE standard

Grimm et al. (2014)


RIFCON eVole Population Model Version 3.0 – TRACE document

TRACE document

Date: June 26, 2018

This is a TRACE document ("TRAnsparent and Comprehensive model Evaludation") which provides supporting evidence that our model:

eVole - RIFCON Common Vole Population Model (Version 3.0)

was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose.

The rationale of this document follows:

Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V. 2010. Ecological modelling supporting environmental decision making: a strategy for the future. *Trends in Ecology and Evolution* 25: 479-486.

and is using the updated standard terminology and document structure proposed in:

Grimm, V., Augusiak, J., Focks, A., Frank, B., Gabsi, F., Johnston, A.S.A., Kulakowska, K., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Schmolke, A., Thorbek, P., Railsback, S.F. 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecological Modelling* 280: 129-139.

and

Augusiak, J., Van den Brink, P.J., Grimm, V. 2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecological Modelling* 280: 117-128.

Copyright © 2018 RIFCON GmbH
www.rifcon.de
info@rifcon.de

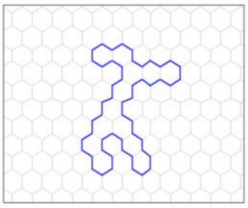
RIFCON (2018)

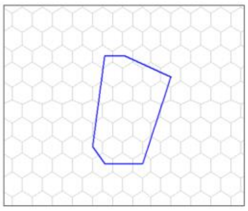
Model mechanisms

8

- Maturation pup -> subadult -> adult in a pre-set time
- Daily background mortality
- Subadults and adults compete for home ranges
- Reproduction when f and m home ranges overlap in breeding season
- Dispersal if lack of food or mating partners

Population dynamics emerge from home range dynamics





Home ranges (RIFCON 2018)

States and processes of repoduction modelled in eVole (start -> mate -> pregnant -> gestation period -> give birth -> lactating -> lactation period -> end)

- Figure removed due to copyright restrictions -

Fig. 1 from 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.

Reproduction (Wang 2013)

Model parameters

9

Life history parameters

Mostly from random distribution at birth

Mortality: daily mortality, maximum age

Reproduction: sex, time to maturity, gestation length, litter size, lactation length, time lag before fertile again

Home ranges: max. number of cells added, food demand (depends on age + sex)

Dispersal: food threshold, maximum distance

Environmental parameters

Landscape: vegetation types (arable field, grassland, hedge), spatial arrangement, seasonally varying forage values (for each vegetation type), seasonally varying vegetation cover, min. vegetation cover for home range

Breeding season

Time and location of PPP exposure, concentration

Coupling to individual-level effects

Main processes 10

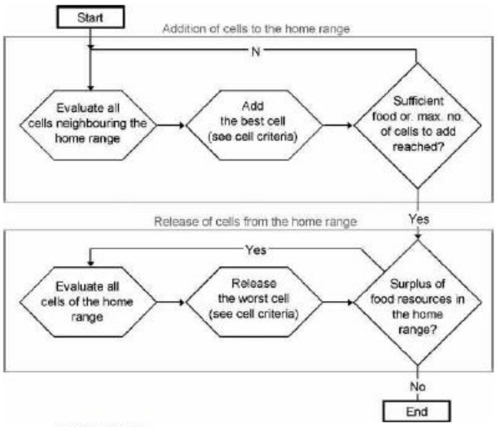
Home range optimization

Adjust food, minimize overlap, sufficient vegetation cover

Reproduction depends on

Home range availability (forage, vegetation cover, population density)

Density regulation based on reproduction



RIFCON (2018)

Home ranges in spring and summer (RIFCON 2018)

Main processes 11

Concerns

- **No species interactions**

  ⇨ Predictions only for isolated populations (no predators + competitors)

- **Few additional environmental stressors**

  ⇨ Parameterization e. g. without additional pesticides

Sensitivity analysis

12

Model Output	β^i	R^2	Parameter
Population density	−0.3697	0.313	Breeding start of females
	0.2225	0.139	Breeding end of females
	0.2024	0.135	Breeding end of males
	−0.1715	0.082	Gestation length
	−0.1530	0.064	Mortality of adult males
Number of offspring	−0.4955	0.401	Breeding start of females
	0.3078	0.193	Breeding end of females
	−0.3050	0.169	Gestation length
	0.2746	0.185	Breeding end of males
	−0.1591	0.057	Breeding start of males
Mean monthly growth rate	0.2496	0.125	Breeding end of females
			ales
			nales
			es
Fem			f the landscape
Mal			f the landscape

Sensitivity analysis: Time to recovery from acute mortality in April and from reduced litter size for two weeks in June. Recovery time increases with additionally induced mortality from 1 month at 10 % mortality to > 12 months at 70 % mortality. Recovery time from reduced litter size increases from 1 month at 10 % reduction to > 12 motnhs at >= 20 % reduction.

- Figure removed due to copyright restrictions -

Fig. 7 and 8 from 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 (2013)

Validation of basic population model

13

(Wang 2013)

Comparison of modelled and observed population cycles across years. Shortening the length of the reproduction season in summer decreases yearly fluctuations in the mean modelled population density in August. A similar decrease in fluctuation of population size in August was observed in field data from increasing lattitude with shorter summers.

- Figure removed due to copyright restrictions -

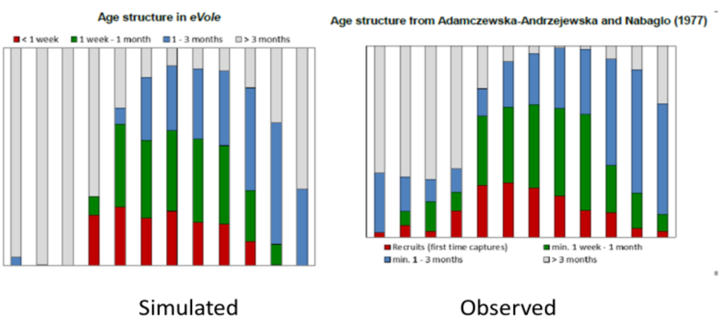
Fig. 6 from Wang, M. (2013). From home range dynamics to population cycles: Validation and realism of a common vole population model for pesticide risk assessment. Integrated Environmental Assessment and Management 9(2): 294-307.

(RIFCON 2018)

Comparison to field studies

Magnitude of population cycles vs. Length of breeding season reproduced

Seasonal variation in age structure reproduced



Model testing

14

No validation of predictions relevant for risk assessment

MODELINK workshop

- Application on hypothetical fungicide

  eVole

  ALMaSS (field vole)

  Liu et al. (2013) (wood mouse)

- Assumed effects on survival, litter size, maturation
- Probability distribution for effects to simulate variation in exposure and susceptibility
- All models suggested low risk due t fast recovery

Predicted effects on common vole populations.

Deviation in the population size from the control populations is clearly visible and larger than 5 % only for the toxic standard exposure scenario (20 % acute mortality of exposed individuals) and also in case of 10 x the expected application of the modelled pesticide.

- Figure removed due to copyright restrictions -

Fig. 5 from Schmitt, W., et al. (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 et al. (2015)

Case study

15

Bastiansen F, Meli M (2017)

Population modelling for the common vole to assess the potential effects following the application of folpet in vines

RIFCON GmbH

Sponsor: ADAMA Makhteshim Ltd

Background

- „Meldoy Combi“ (fungicides folpet + iprovalicarb) for spraying in vineyards
- Chronic risk assessment of folpet: reduced maternal body weight and reduced litter size, delayed long bone epiphysis ossification in rabbits

Background

- Simulation study to demonstrate that expected chronic mortality will not result in long-term population effects

Case study

16

Environmental scenario

Landscape

- 75 % vine yard (exposed) + 25 % grassland
- Mowing 15th May + 15th July

Exposure

- 10 x 1.5 kg/ha, 7 d interval, in May – July or July - September
- 100 % of pesticide reach ground
- DT50 = 6.22 d
- 5 y pre-treatment, 10 y treatment, 10 recovery

Starting conditions

- 1st January,
- 2 weeks pre-run to establish home ranges
- Mean population density from 15 y test run

Parameterization for 3 generations per year

Bastiansen & Meli (2016)

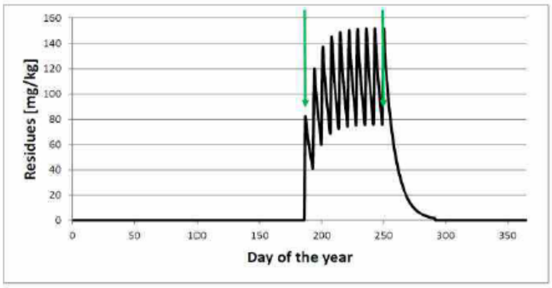


Figure 13 Residues on plants after late application of folpet (nominal application rate)

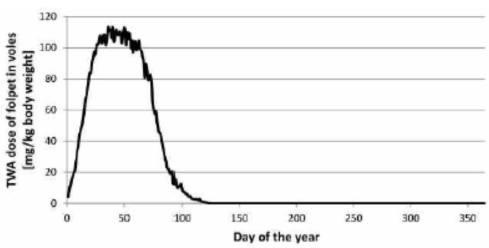
Case study

17

Individual-level effects

Individual exposure

- DDD following EFSA guidelines, considering body weight (not in model), food consumption, proportion of contaminated food, residues in food, dissipation rate



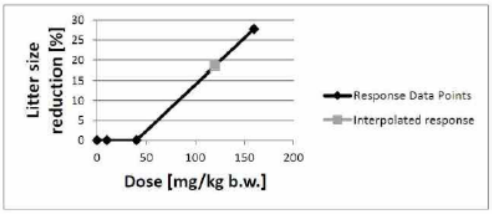
Mean TWA dose of folpet in exposed individuals after 10 applications between January 1st and March 5th (Bastiansen & Meli 2016)

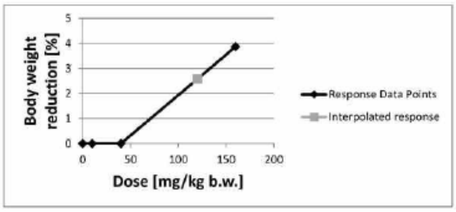
Case study 18

Individual-level effects

Effects observed in rabbits (Rubin 1986) and implementation for voles

- 10 mg/kg/d DDD: NOAEL
- 40 mg/kg/d: reduced maternal body weight (-> translated to **mortality during gestation** using dose-response from Oksanen et al. 2007)
- 160 mg/kg/d: increased abortion risk (ignored)
  additional rib more common (ignored)
  temporarily delayed tail bone ossification (ignored)
  reduced long bone epihpysis ossification (lethal-> reduced **litter size**)
  reduced newborn body weight (-> increased **juvenile mortality**)

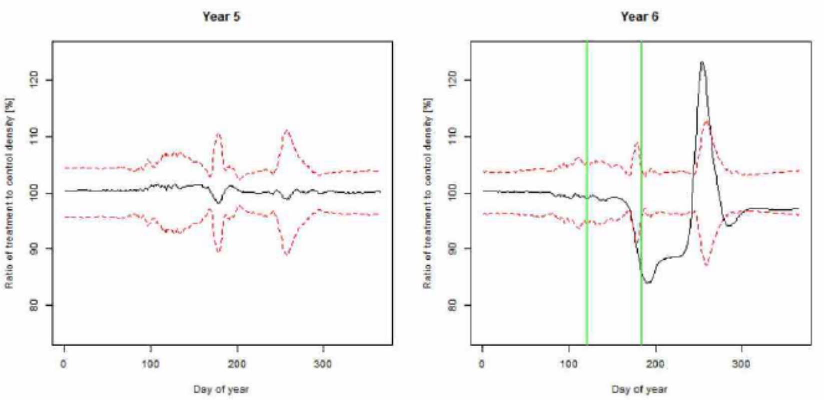




(Bastiansen & Meli 2016)

Case study 19

Modeling results Early application scenario



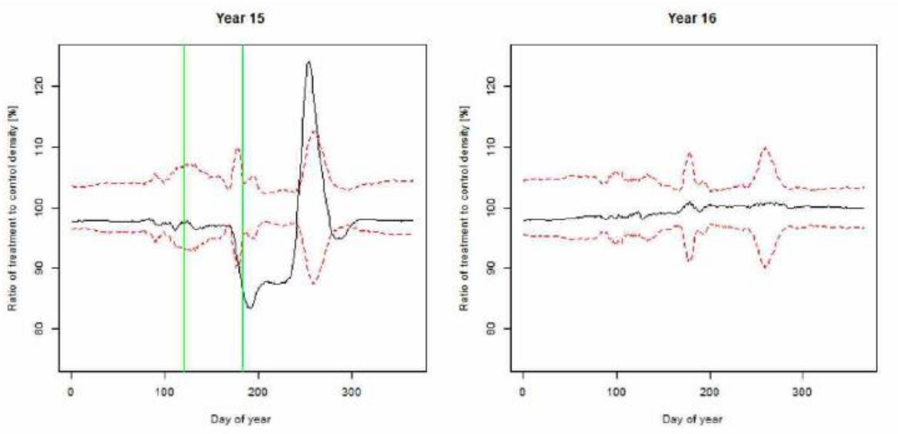
Relative population size before (left) and in 1st year of PPP application (right) (Bastiansen & Meli 2016)

- Populations controlled by forage and vegetation, not by PPP effects
- Sublethal effects stronger than increased mortality

Case study

20

Modeling results **Early application scenario**

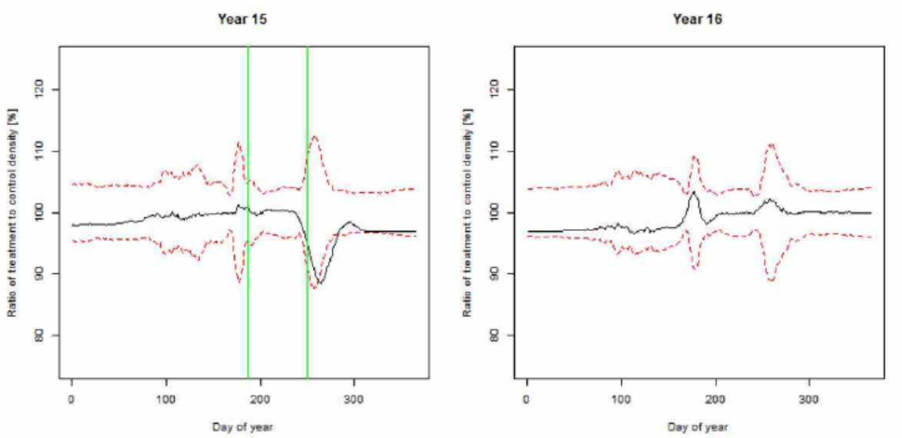


Relative population size in last year of PPP application (left) and thereafter (right) (Bastiansen & Meli 2016)

Case study

21

Modeling results **Late application scenario**



Relative population size in last year of PPP application (left) and thereafter (right) (Bastiansen & Meli 2016)

- Less reproduction after exposure, no overshoot
- Effects slightly stronger (3 % reduced population size at end of year)

Case study 22

Conclusions

Uncertainties leading to overestimation of risk

- 100 % fungicide reach ground
- Voles only feed on diet with higher residues
- Toxicological data probably specific to rabbits, but applied to voles (surrogate species)
- Effect on body weight (maternal, F1) translated to acute mortality (without indication from lab)

Uncertainties leading to underestimation of risk

- No species interactions
- Not all pot. relevant environmental stressors considered (e. g. other pesticides, diseases)
- Data not well suited for dose-response relation
- High uncertainty for individual-level effects
- Landscape composition pot. not realistic worst-case
- No off-field exposure from drift

Risk assessment rather balanced

Summary 23

Strengths

- Parameterization with field data (incl. some natural stressors)

Concerns on model development

- No species interactions (no food web effects)
- Not all pot. relevant environmental stressors considered
- Uncertainties in parameterization

Main concerns on model application

- High uncertainty in the implementation of individual-level effects
- Model not validated before application
- No specifically defined protection goal available

Jeremias Becker, Tjalling Jager, Stephanie Kramer-Schadt, Mathias Franz, Matthias Liess, Magali Solé, Sabine Duquesne, Silvia Pieper, Steffen Matezki, Tobias Frische, Jörn Wogram

Thank you!

Umwelt Bundesamt

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

DEBtox Research

Leibniz-Institut für Zoo- und Wildtierforschung

Leibniz-Gemeinschaft

Annex 2: Minutes from the Symposium

Minutes from the symposium "Evaluation of toxicological and ecological effect models for risk assessment of plant protection products", organized by UFZ and UBA in Berlin, 19th and 20th of September 2019, as part of the UFO Plan project 3715 67 408 0.

Context and Objectives of the Symposium

This symposium was organized by the **Helmholtz Centre for Environmental Research** (Helmholtz Zentrum für Umweltforschung UFZ, Germany) and the **Federal Environmental Agency**, Germany (Umweltbundesamt UBA).

The intention of this symposium was to disseminate the main outcomes of the UBA research project UFOPLAN 3715674080. This 3-years project (started in 2016) aimed at reviewing critically existing mechanistic and ecological models potentially suitable for the refined risk assessment of PPP.

During the symposium, the evaluation of toxicological and ecological effect models for risk assessment was presented and discussed from a scientific and a regulatory point of views. They were based on case studies delivered as part of dossiers to the UBA.

The symposium also aimed at providing a platform for risk assessors, representatives from academy and industry / consultants as well as from the EFSA to discuss further potential developments necessary for a future successful implementation of ecological models in risk assessment.

The symposium took place in Berlin in September 2019 with 35 participants, divided as follows: 20 participants from regulatory authorities of 11 Member States, 3 participants from academia and 12 participants from industry/ consultant as well as 1 representant from EFSA. In addition, 4 members of the consortium participated. The Organising Committee included the members of the consortium and risk assessors from the department of Plant Protection Products in UBA.

The Minutes of this symposium are presented below. Please note that most inputs received from the participants during the commenting round of the draft Minutes were implemented in the Minutes either to alter the text or as "post-symposium note" within the corresponding sections.

1 Session on Individual Level Models

The discussions were based on case studies presented by Dr. Tjalling Jager (see Annex 1).

1.1 TK-TD Modelling, GUTS

Applying the modelling cycle

The Sci. Op. on Good Modelling Practice (GMP) (EFSA, 2014) says that Model development should follow the modelling cycle with all steps mentioned there (see GMP, section, p13).

However, the presenter stated that there is no need to start the whole "modelling cycle" from scratch for each new GUTS question. He suggested that a simplified version of the modelling cycle might be in some cases more appropriate because not each step can be applied to each model. E.g., uncertainty analyses are not applicable to the GUTS approach per se (basically because no parameter values have been built in but must be fitted to the data). Therefore, uncertainty analysis makes only sense for a specific model application of GUTS.

Some participants proposed that while evaluating a model, concepts (i.e. entities, algorithms, equations and the links between them) should be separated from software implementations and model applications (because e. g. for the concept of GUTS different implementations are available).

In general, the model development according to the modelling cycle should be performed based on one example independent from its application for a specific chemical-species combination, whereas further applications of the model might use a simplified modelling cycle. This means that for the GUTS model, each application (i.e. specific chemical-species combination or even a specific data set) can have a simplified evaluation, in addition to the general evaluation of the model. In principle, the majority of the participants agreed.

It was further pointed out that this approach has been followed for GUTS presented in the Sci. Op. on TKTD Modelling (EFSA PPR 2018).

As a summary, it can be stated that a simplified version of the modelling cycle might be in some cases sufficient, e.g.:

- when not every step of the modelling cycle is applicable (e.g. uncertainty analysis not applicable for GUTS in general but only for a specific application).
- when the whole modelling cycle was already run during the model development. E.g., if a conceptual model and its implementation were assessed and considered as satisfactory according to the modelling cycle, it is then not necessary to re-assess it for every application, i.e. only the regulatory model including the environmental scenario (i.e. the exposure scenario for GUTS) and parameterization will need to be assessed in the context of the new application.

Potential of models to identify risks

The potential of models to identify more risks than risk assessment based on experimental work was discussed. The presenter pointed out that in principle models could in some cases identify more risks than in the experiment. For instance, with the GUTS model it is possible to virtually prolong the duration of an acute fish test to more than 4 days (i.e. extrapolate to effects for longer exposure durations). Thus, e.g. in case of an active substance (a.s.) with an irreversible mode of action (MoA), the prediction of an LC_{50} with GUTS might be more accurate than the one calculated with a classical dose-response analysis (e.g. Probit regression), as the data for all observation points are used in the model fit. Furthermore, the model fit can indicate whether the LC_{50} is expected to decrease further after the test duration (see slide 4).

It was pointed out that such effects are usually covered by the tests additionally performed for the chronic RA. The presenter disagreed that these effects might be covered, because experimental conditions for acute and chronic tests can differ; e.g. in comparison to the conditions in acute tests, in chronic fish tests small fishes are properly fed and are growing. Therefore, the growth might dominate the kinetics and might also lower the internal concentration because of dilution. As a result, the effects might not be triggered.

In this context, a participant questioned the relevance of modelling an acute LC_{50} for more than 100 days (see slide 27). Indeed, the duration of the lab tests is very important to detect effects which need some time to become visible; the quality of the GUTS parameters strongly depends on the duration and on the amplitude of the effects used for parameterization (e.g. 5 or 50 %). It was agreed that the conventional (current) acute RA is usually based on test durations of e.g. 4 days for fish, but in some cases, it might be appropriate to check the mortality for longer time scale.

To conclude, participants agreed that using GUTS is not a matter of acute or chronic RA, but rather a matter of how to consider lethal effects (i.e. having the possibility to consider lethal effects on different time scales than experimentally available).

Comparison of GUTS and classical dose-response

It was discussed if GUTS is a better predictor than a “classically” derived ECx/LCx and if it improves the Tier 1 RA?

Some participants raised the issue on how to be sure that the GUTS model will deliver better predictions than the classical dose-response analysis used in standard tests since in case the decision would rely on GUTS, then risk assessors want to be sure that the correct decision is taken.

The presenter answered that GUTS might lead to a higher accuracy in the prediction of the LC_{50} because the model (i) makes use of a well-established theory on the mechanisms in TKTD instead of being purely descriptive, (ii) can integrate all information available from an acute standard test (see point below) and (iii) can be prolonged to simulate an extended test duration (see point above).

This was acknowledged. Also, it was added that confidence intervals in GUTS generally correspond to those of classical approaches for the LC_{50}, but are much smaller for the LC_{10}.

Use of GUTS models in current scheme

GUTS models can be used in different ways in risk assessment.

Use as Tier 2C approach: They are frequently submitted as Tier 2C refined assessment during product authorization for assessing effects of refined exposure profiles, but it was raised by some participants that the protectiveness of the Tier 2C approach is questionable. However, this issue applies both to modelling-based as well as experiment-based studies in the Tier 2C approach (see section 4.5 in EFSA Supporting publication 2019:EN-1673).

Calibration of the current RA scheme: It was acknowledged that the validation of models is clearly a different issue and should be differentiated from the validation of the RA scheme. However, it was also stressed that the current RA scheme is not validated with field data (only restricted validation for some aquatic cases with surrogate field data, i.e. mesocosm data); therefore, any type of refinement could lack of protectiveness. For the refinement using the Tier 2C approach (modelled-based or experimental-based), this adds to the critical points mentioned above.

The potential use of GUTS as module for individual-level effects in population models was discussed later (see below).

Cost/benefits, expert judgement and training

Some participants (regulators) raised the issue on the cost/benefits of using such a model. It was explained that usually a "single" risk assessor has to evaluate a whole dossier within a defined time-scale (i.e. risk assessment for all groups of organisms at all tiers including refinements; in some Member States (MS) the assessor has even to evaluate the fate part). Therefore, the issue of cost/benefits is of utmost importance.

Some risk assessors were interested in knowing "what kind and how much training for assessing models would be then necessary?"

Modelers answered that the effort and resources to be invested depend on how the model will be used and on the process in focus, e.g.:

Evaluating the calibration or the validation of a GUTS model for a new species or a new a.s. requires some effort. However, a validation might not be needed for cases where only some types of extrapolations are done, e.g. using the model to predict effects after longer durations than those used in Tier 1 tests or for substances having irreversible MoA.

It was also raised that fitting the model should not be a problem, but e.g. how to interpret the fit needs some expertise.

One participant (modeller) raised the issue that using different software packages without guidance requires more expert judgment; this is critical when focusing e.g. on LC_{10} or LC_{50} values.

Linking exposure to effects

GUTS models enable to link exposure to effects on survival. Aspects of the discussion are summarized below:

Better use of the data

Some risk assessors asked how can inputs with low resolution (LC_{50} and PEC) result in high resolution in the model. It was answered that GUTS is making a better use of the data than the classical calculation of the LC_{50} because it considers all information (raw data) on exposure and effects recorded for every time point. In GUTS the effects due to exposure are estimated in a better way since variations in magnitude and duration of exposure (e. g. data from different observation times after the beginning of a test) can be used as input in a single model.

Thus, to conclude an appropriately calibrated and validated model that predicts effects over time (extrapolation for longer periods than tested experimentally) can be more accurate than the current method deriving an endpoint only for one specific exposure duration.

Allow for simulating other exposure conditions

The participants were of the opinion that, once the model is calibrated and validated with appropriate lab data, it can then be used to simulate more exposure scenarios than what is experimentally possible. This is actually the position of the EFSA Sc Opinion.

Tier 2C experiments usually test only one exposure pattern, e.g. one or few peaks during a standard test duration (acute or chronic). However, FOCUS exposure profiles last longer or might have a different exposure pattern than the one tested in the refined exposure test. Thus, a fit-for-purpose GUTS model (i.e. properly calibrated and validated) can help to extrapolate results of such a study to other FOCUS profile situations.

Better understanding of the underlying mechanisms

Standard Tier 1 tests are designed to be conducted under standard/constant exposure; the GUTS models (once appropriately calibrated and validated) will allow to simulate different (non-constant) exposure conditions (see point above). Thus, some participants were of the opinion that by varying the exposure regime, GUTS might help to better understand the mechanisms underlying the effects. E.g., it might be studied whether effects are triggered by a short exposure such as a peak concentration, or rather by a longer exposure duration such as a time weighted average concentration). In a context based on models, it was also noted that a strict separation between acute and chronic is not meaningful anymore, instead the separation should clearly be between lethal and sublethal effects.

Needs for GUTS validation

One participant raised that the EFSA Sci. Op. on TKTD modelling describes how the validation has to be performed in case a GUTS model should be used for Tier 2C. However, in this document the criteria for validation of vertebrate models are less strictly defined, since generally vertebrate testing should be reduced.

EFSA observes a trend in increased submission of refined exposure tests (Tier 2C) at the EU level (Peer Review), especially for fish. It was suggested that the possibility of using these tests for validation of the GUTS model should be further explored.

1.2 TK-TD Modelling, DEBtox

What is a DEBtox model, when using DEBtox?

In a first part, it was explained that DEBtox is a special model that belongs to the "DEB-Model Family", which regroups any kind of models based on the energy budget theory. The energy budget theory assumes that there is a trade-off at the individual level between growth, development and reproduction. A DEBtox model is then a special case of a DEB model applied to toxicant stress. These models are thus one possible approach for addressing sub-lethal effects for individuals.

In a second part, a case study was presented. However, this case study does not represent a classical application of DEBtox but may rather be considered as a "standard DEB" model applied to an ELS (early life stage) toxicity study. This was discussed and agreed after a participant involved in the development of the model for this case study clarified the aim: The use of this model was to explain why no effects were observed at the highest concentration tested in the refined exposure ELS, although such a concentration would have trigged mortality in a Tier 1 test (in a test performed under constant/ standard exposure conditions, i.e. with concentrations maintained for the entire duration of the test).

A risk assessor raised the point that the focus should be on "effects" rather than on "no effects". This was acknowledged by the presenter, but he mentioned that cases are mostly available either for experiments performed at concentrations that do not trigger effects or for experiments that are too short to demonstrate effects.

Tests with "no effects" are an issue when (i) a No-Effect-Threshold has to be defined, and (ii) used to validate a model. Clear effects are needed to adequately parameterize models; indeed, tests focusing on NOEC often show no or too small effects.

To gain confidence in a model, it might be useful to demonstrate that a model is able to predict effects for conditions where effects are expected, although the focus of the RA is generally on 'no (unacceptable) effect'. This could be considered as part of the model testing.

The database Add-my-Pet (AmP)

Post-Symposium Note

Both the DEBtool and AmPtool are freely available and open source (https://github.com/add-my-pet). These programs are used for preparing the AmP entries and performing comparisons and statistics on the AmP database. In addition, all AmP entry code is fully available together with the data used to perform the calibration, the reference to these data, and any point of discussion that the authors might have had: https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/

In the case study, for the parameters related to the rainbow trout, the authors used information from the AmP portal. The pros and cons of this approach were discussed.
As an advantage, it was stressed that using entries from the AmP portal might be useful especially when a "full DEBtox model" is used because in that case the model requires more data than typically derived from toxicity tests.

On the other hand, it was stressed there is no real "quality control" on the single entries of the database. Indeed, even if a kind of quality control exists, a single entry relies more on expert judgement (see: http://www.debtheory.org/wiki/index.php?title=Completeness). Using this database can be linked to some uncertainties. Thus, a more rigid peer review may be necessary if the database was to be used as a standard source of information for ERA applications.

Another issue with using AmP is that adjustments between the parameters and the measurements in the lab tests might be needed to match the situation in the test. This was illustrated in the case study presented (food level tuned to match final body size in the test). In principle, such adjustments might be acceptable, if supported by data.

How to improve DEBtox models

It was acknowledged that appropriate data are needed to get a good model. In that context, a participant asked about recommendations or suggestions for better data generation. How can experimental testing be improved to get better data for model development?

It was proposed that more intermediate measurements would be needed (i.e. not only at the end of the test) and ideally all relevant parameters should be measured. This is especially true for the fish, for which modelers seem to agree that there is a deficit of knowledge about the processes involved for the early life stages (ELS), i.e. what is happening in the eggs, what are the toxicokinetics of this early stage etc. For daphnids, the data availability seems to be better.

Similar statements regarding the improvement of data recording during the test to increase the predictive power of models also apply for GUTS.

Decision making

The questions raised were if DEBTox is a better predictor than the current approach in ERA and if it improves the RA.

A participant (risk assessor) raised the issue on cost/benefits of using models in the RA: With their limited resources, risk assessors want to achieve the best results for the environment. So it was questioned if they can reach that with models, i.e. if they could make completely different decisions or if they dissipate their resources.

The main issue is whether risk assessment can lead to better regulatory decisions (i.e. with less uncertainties) when using new tools in risk assessment, compared to the current approaches i.e. if the models are improving the performance of the RA?

Post-Symposium Note

The role of risk assessors is to provide the most accurate description of the risk as possible, so that the risk manager can take decision in order to find the optimal balance between food production and environmental protection. If it can be proven that new tools can better describe the risk (with less uncertainties), then risk assessors shall use them. DEBtox, like GUTS, can enable extrapolations to non-tested situations and to make use of more data available from tests, as compared to point estimates such as the classical ECx approach.

Potential applications

In principle, and in-line with the EFSA Sci. Op. on TKTD Modelling (EFSA PPR 2018), the potential of the DEBtox models to assess sub-lethal effects was acknowledged.

A risk assessor raised the possibility of using DEBtox models to better understand the underlying mechanisms observed for some endpoints and/or to answer regulatory questions that cannot be solved currently. This was illustrated with an example on earthworms, for which an increase in body weight (bw) is sometimes observed in toxicological tests for some substances. Currently such effects are not always considered relevant to set the endpoint. However, according to the DEB theory changes in bw indicate changes in the energy budget of the animal with potential effects on some other traits.

The presenter answered that in principle this may be possible, but it depends on whether we find a model that makes certain regulatory questions easier to answer. Modelers agreed that e.g. this issue on bw gain in earthworms might fit to the purpose of DEBtox models. However, they stressed the difficulty of handling data with soil organisms especially due to (i) their different routes of exposures (contact and oral) and (ii) the complex behaviour of the a.s. and the organisms in the soil profile.

To conclude the session, the same participant stressed that currently we have to evaluate what is delivered in dossiers and suggested that we change that, i.e. the use of models should be more driven by the questions we have and should thus be more oriented / guided by risk assessors. Please note that this applies to all kind of models and not only to DEBtox.

Post-Symposium Note

A difference between GUTS and DEB with respect to data / parameters needed is that DEB models need several data just to describe the control organisms; this is not the case in GUTS for which just background mortality is needed.
The issue on more collaborations between risk assessors and modelers in the development of models was generally acknowledged by all stakeholders (at the symposium and during the commenting phase of the minutes).

There might be some other potential applications for DEBtox models, however they are not yet explored or submitted.

2 Session on Population-Level Models

The discussions were based on case studies.

2.1. General introduction

An introduction on Individual Based Modelling was presented by Dr Stephanie Kramer-Schadt.

2.2. The IBM *Chaoborus* Population Model

The Chaoborus IBM was presented as an example of population model by Jeremias Becker (see Annex 1).

General issues on this case study

The Chaoborus model was illustrated by a case study on beta-cyfluthrin. This case study was recently submitted in Germany in the frame of the renewal of the active substance for which DE was RMS at the EU level.

The aim of the model was to show the ability of the population to recover at the threshold concentration (ETO), used for RA. Moreover, the model can be used to explore the interplay of dynamic exposure, effect modelling and population modelling which allows a different, dynamic view, as these influence the intensity and duration of the effects.

The model developers of the Chaoborus model pointed out the existence of another application of the same basic / ecological model (submitted for the active substance alpha-Cypermethrin, RMS Belgium) which was already reviewed at the EU level and of better quality since validated with three independent datasets from mesocosm studies. They were of the opinion that i) the model itself should not be judged based on a model application with a relatively small data basis, ii) evaluation of the model and its application should be clearly separated, and iii) if a model cannot be applied to a specific case, this does not necessarily mean that the model per se is bad. However, the approach foreseen in Appendix B of the EFSA Sci. Op. on GMP (2014) "Summary checklist for model evaluation by the risk assessor" is about the evaluation of the formal (basic / physiological) model as well as its suitability for regulatory purposes (with the latter corresponding to the model application).

The consortium[1] and the UBA answered that there was no selection of "good" or "bad" applications / case studies in the context of this project. The only limiting criteria was that DE was rapporteur member state (RMS) in case of EU-active substance evaluation or zonal rapporteur member state (zRMS) in case of PPP authorization, in order to guarantee full access to the original data. Moreover, it can be assumed that if a model is submitted to the authorities, then it is considered suitable (i.e. fit for purpose) by the applicant. The case studies evaluated in the current project only reflect the types of models submitted.

Complexity of models. What do risk assessors require?

It was acknowledged that the evaluation of models might be difficult for non-modelers.
Then, the discussion was about the degree of complexity needed in a model.

Some participants were of the opinion that "simple" models with more conservative scenarios might be more appropriate, whereas other participants were of the opinion that models need to be realistic.

[1] The term "consortium" in this document only refers to the project group and not to the participants of the symposium.

A model needs to include the relevant processes for delivering suitable output that addresses the specific risk assessment question of the model application; thus, they do not necessarily have to be very complex but oversimplification should be avoided.

The consortium pointed out that the risk assessment needs “balanced models”. This means that at least the most relevant factors must be included. For instance, if the species interaction is relevant for describing the environmental risk, then it needs to be integrated independently of the degree of complexity required.

Post-Symposium Note

Some participants commented that if the model is parameterized with mesocosm data, then parameters such as growth rates are already implicitly including effects of interactions (intra and interspecific), and thus effects of community would not need to be explicitly implemented. We generally agree. However, we are of the opinion that i) this is insufficient since the potential effects of stress due to species interactions is modified under exposure conditions; therefore, as long as the model is only tested against control mesocosm, this would remain insufficient, and ii) the environmental conditions in such field or mesocosm studies need to be representative for the conditions that the simulated population will experience in the field. The development of individuals in the *Chaoborus* IBM has been parameterized mainly with data from laboratory studies and with data from highly artificial microcosms (covered by nets, protected against predatory or competing species) which we do not consider representative for the influence of a real community in the field.

Level of conservatism and environmental scenarios

A modeler suggested that assessing if a model is conservative enough depends on the way the model is used, especially which environmental scenarios were run with that model.

Post-Symposium Note

“environmental scenarios” refer to “a combination of abiotic, biotic and agronomic parameters to provide a realistic worst-case situation”, as defined in the EFSA Sci. Op. on GMP (2014).

There was a common agreement that environmental scenarios are of main importance because they are directly related to the outputs of the models. Further, some modelers were of the opinion that there is a need to agree on some standard environmental scenarios (e.g. appropriate food level, temperature, landscape structure, competition…). Indeed, using a different environmental scenario might change the outcome of the model and require a new evaluation.

Thus, instead of being concerned whether a model is conservative enough, some participants proposed to rather use the model to rank various situations in terms of risk and margins of safety. For instance, with regard to the GAP, simulation studies models may investigate how much an application rate could be increased until the models show an effect? This would be similar to the approach proposed for GUTS in the EFSA opinion on TKTD (e.g. LPx/EPX).

Post-Symposium Note

Studying margins of safety is not an alternative to evaluating the conservatism of a model application which results from the selection and implementation of processes that potentially increase or decrease the real risk. A model that contains only mechanisms that potentially decrease the risk will predict higher margins of safety than a model that contains only mechanisms that potentially increase the risk.

To conclude, participants agreed that exploring the environmental scenarios (for different pesticide properties and uses as in Fig 5 of the Sc Op on GMP) and their representation of realistic worst-case situations should be critically considered and assessed when evaluating the conservatism of a model application.

1.1.1.1 Exploring environmental scenarios and/ or GAP

In principle, most of the participants, especially the risk assessors, are in favour of using models to achieve a better description of the risk (e.g. under which environmental scenarios are lower or higher risks identified in the field), instead of having models showing "no risk/ acceptable effects". Indeed, the latter kind of applications is mostly delivered in the current dossiers (i.e. using refinements to demonstrate a safe use/ waive an unacceptable risk identified at lower tier).

However, one participant mentioned that demonstrating "no unacceptable effect in the field" is the aim of the current risk assessment as mentioned in the legislation and is thus valuable. But it would be also relevant to demonstrate the exposure level at which effects become pronounced. A modeler acknowledged that one could get suspicious towards the modelling approach, if model calculations are delivered only in dossiers for showing no risk. However, modelling studies showing a high risk (e.g. population at risk extinct) exist but they are usually not included in the dossier by the applicant and thus not submitted to the authorities (or would also not appear in EFSA's EU level evaluations of active substances). Thus, the acceptance of the modelling approach can be difficult if the risk assessors are only delivered those models and model calculations showing acceptable risk.

However, this participant confirmed that models are useful to identify where the risk is.
For example, models might be useful to:

- Identify the best possible application windows, i.e. which application window would lead to least risk with regard to the most sensitive life stage?
- Extrapolate to predict effects of exposure patterns different than the one tested in a mesocosm study on a particular species.
- Extrapolate to assess environmental scenarios that cannot be assessed experimentally.
  In this context, the authors of the Chaoborus model raised that by switching-off the immigration, a worst-case scenario such as "isolated pond" could be simulated in the model, which is practically not feasible in semi-field study.

Post-Symposium Note

During the commenting phase some participants proposed two options for exploring the environmental scenarios: i) either the applicant has to additionally integrate simulations for unacceptable scenarios into his dossiers, or ii) the authorities carry out the modelling and scenario selection by themselves. Note that the latter is recurring under "Exploring the intended uses / GAP" in the section below on spatially explicit IBM.

Uncertainties and need for criteria for evaluation

This part of the discussion is related to slide 12 of the presentation of the model on *Chaoborus* (Figure 8 in Strauss et al. 2016). Based on the simulations presented, a risk assessor noted that the uncertainties around the predictions appeared at first glance quite "variable" and "wide". This is especially true for the L2 larvae. But it was also noted, that the model uncertainty seems to capture well the variability observed in the experimental data (also large) with the existence of a time discrepancy between the modelled and the experimental data). Thus, the issue of evaluation and decision-making based on simulated data showing such patterns was raised.

It was answered that such uncertainties are difficult to aggregate into one value since part of the variations observed might come from natural variability. Multiple patterns (i.e. combination of different patterns) might have been captured at the same time (processes showing e.g. responses to pesticides, natural variability, interactions between populations of individuals from a same size class, interactions between populations of individuals of different size classes).

It was further stressed by risk assessors and EFSA that for the case where they would be confronted with the evaluation of such patterns of variation, criteria would be needed to ensure a correct and harmonized decision.

A risk assessor also asked about the general uncertainties linked to the use of models.

Post-Symposium Note

About this last comment, please consider the "Summary checklist for model evaluation by the risk assessor" presented in Appendix B of the EFSA Sci. Op. on GMP (2014).

1.1.1.2 Species interactions: population versus community models

Post-Symposium Clarification

Please note that the concerns highlighted in this section relate to the use of models at higher tier (i.e. Tier 3 of the aquatic risk assessment).

One participant from the consortium stated that species interactions usually lead to higher vulnerability of the species under focus. However, some participants replied that this should not be a general assumption/rule because there might be situations where it is the other way around. E.g. in a prey / predator system, if the predator is affected (more than the prey), it is possible that the prey recovers faster. This could become of regulatory relevance but only in a context where the recovery option is considered.

The developers of the Chaoborus model explained that the basic / ecological model was parametrized with data from laboratory studies and tested / validated with mesocosms. Due to the good match between model and mesocosm experiments, they concluded that the lack of interspecies interaction in this specific model is not relevant for field populations, i.e. the growth rate data implemented in the model is realistic and already account for/ include species interactions. According to another participant this approach is sufficient to address the species interactions (density-dependence), since it is already included in the parametrization. Other participants had different opinions:

- since no predators are included in the Chaoborus model, such a population model can thus not predict situations at the community level, or
- since Chaoborus has no predator in the field, considering predatory interactions may therefore not be relevant in this case.

Please note that cannibalism is implemented in the model.

Post-Symposium Clarifications

The consortium understands that the Chaoborus population model was (partially) parameterized with physiological data (e.g. growth rate) from laboratory studies without interacting species. Thus, potential effects of interacting species on growth and development were covered neither explicitly (through the simulation of interacting species) nor implicitly (through parameterization with physiological data from populations that interact with other species) in the model. However, it is likely that in a natural community Chaoborus is exposed to antagonistic species such as amphibians and large invertebrates who may act as predators and competitors. Effects of interacting species on growth and development may decrease the potential of a model species to recover; without implicitly or explicitly covering the effects from species interactions a population model may therefore underestimate the risk of pesticides. Additionally, pesticides may change species interactions (due to the fact that different species will be affected in various extents); this potentially increases the community-induced delay of population recovery further if sublethal pesticide effects are stronger for the model species than for its prey / competitor. The community effects on population recovery outlined above may be observed in well-designed mesocosm studies. As mentioned in the section above ("Complexity of models"), if species interactions are a potentially relevant factor, then they need to be integrated also in population models in order to increase realism in risk assessment.

In the commenting phase, a participant wrote that "this clarification suggests that explicit interactions with toxicant feedback (e.g. multiple species IBM) should be included in every population model. However, this would then represent a community model and not a population model anymore and it is important to make the distinction between both since they have their own interest and specific aims. Community models include more interactions but are more complex to build, calibrate, validate, and the interpretation of the results can be difficult."

An important aspect was raised by a risk assessor for clarification: population models are frequently presented as single population models but the protection goal is the population in the field, i.e. the population within its community. Therefore, in principle a single population model that considers the species most at risk outside the context of its community cannot be considered as appropriate. Indeed, considering the community context is relevant not only for determining parameters such as population growth rate but also for recovery processes.

Post-Symposium Clarification

Please note that in the report of EFSA PPR (2019) on general recurring issues, it is written: "The 'population experiment' mentioned as a Tier 3 in the tiered approach, means that the focus is on a specific population within a community"; although this was stated in the context of experimental approaches, it also applies to modelling approaches; indeed the Tier 3 as in the Aquatic GD (EFSA PPR 2013) refers to "population and community level experiments and models" (Figure 1). Also please note that if a risk assessment is based on experiments which do not always include a "community context", assessment factors are used to represent this fact in order to tackle the protection goal considering „population in the field".

In addition to the above elements, a modeler mentioned that migration (aerial dispersal) may be more important than growth rate in the Chaoborus model; the focus should thus be on such important parameters even so they are hardly quantifiable or poorly investigated. Such information should be communicated to risk assessors.

Ecological realism

While discussing the use of the Chaoborus model to address potential recovery of this species in the mesocosm experiment, a modeler mentioned that some years ago the EFSA published a document in which it was stated (among others) that the RA should be in some cases more realistic. This participant was of the opinion that such requirement could be provided by using models.

Post-Symposium Notes

In our opinion, the participant is referring to the EFSA Sci. Op. on Recovery in Environmental Risk Assessments (EFSA Scientific Committee 2016). Indeed, this Sci Op. mentions on p. 19 that "if the problem formulation phase of the ERA reveals that recovery of NTOs is an issue for a potential stressor and has to be addressed, a conceptual framework can guide the process to increase the realism in the assessment of ecological recovery of populations of vulnerable NTOs in agricultural landscapes". Beside (semi-)field experiments, the conceptual framework of this approach presented in Figure 4 is proposing "ecological modelling" as an ERA tool for recovery.

However, some participants questioned the ecological realism gained from using population models in ERA, compared to the current methods such as higher tier experimental approaches. This is important in deciding if the ERA should include the models or remain as it is now.

One risk assessor asked about the relevance of discussing the recovery because in most cases the "Effect-Threshold" (ETO-RAC as in the Aquatic Guidance Document, EFSA PPR 2013) is more suitable for RA, especially in agricultural landscapes exposed to multiple applications of various pesticides. But models are often submitted to show that despite unacceptable effects shown when using ETO-RAC, recovery will however occur at this concentration (i.e. risk becomes acceptable).

Post-Symposium Notes

In principle, both threshold and recovery options are currently reported in risk assessment. Also, to be noted is that population models can be used not only for deriving an ERO but also for an ETO (if dealing with observed sublethal effects), since the same processes that are relevant for recovery basically determine as well population dynamics.

Validation

Since the environmental context in which population models run is often related to the field scale, the following minutes try to distinguish aspects of the discussion related to the validation per-se of a population model *vs* the calibration / validation of the whole risk assessment scheme.

Validation of models

In principle, the participants agree that a model cannot be labelled as "valid" or "not valid". For example, it could be valid for a specific scientific question or to rank risk but not for deriving endpoints to be used in risk calculations for the ERA.

On slide 13 the presenter claimed that the validation study on predicted population effects after pesticide exposure was of limited use because the mesocosm data showed no long-term population effects that could have been reproduced by the model. He concluded that the study merely demonstrated that the model is not overly conservative (it does not predict effects that have not been observed), while it would be more interesting for risk assessors to know that the model is sufficiently conservative (it predicts effects that have been observed).

A participant objected that there were actually strong effects observed in some treatments at some time points in the mesocosm study used for validation. The presenter responded that only short-term effects (up to ca. 14 d after the last exposure) were observed; the magnitude of these short-term effects (acute mortality) were known prior to the simulation study and served as input to the population model (parameterization of the GUTS module), not as output. In the opinion of the presenter, the relevant output of the population model that should be validated (due to its intended use for risk assessment) is the magnitude and duration of a long-term population decline.

Calibration / validation of the RA Scheme

The risk assessment scheme should be validated towards the reference tier (i.e. field conditions).

But the current aquatic RA is only calibrated towards a surrogate reference tier (semi-field micro-/ mesocosm studies). There is no validation yet towards the field. Regarding the current terrestrial RA, neither calibration nor validation are fully clear.

It should be noted that such a surrogate reference tier can provide relevant data to develop models. However, a participant raised the issue that a mesocosm study can deliver results showing more or less sensitive effects depending on how it is performed (e.g. time of the year, nutrients available, composition of the community) and the endpoint derived from a mesocosm can thus vary in a significant extent. Therefore, the model developed based on mesocosm data still have uncertainties regarding the predictions of field conditions.

The protectiveness of the current RA scheme is questionable. Indeed, effects are observed in the Field (EPIF workshop, 2005), and concentrations above the RACs are observed for many substances in an ongoing monitoring program of small waterbodies in agricultural landscapes in Germany. These exceedances and effects observed in field need to be analysed in order to know if they are linked to non-respect of the good agricultural practices in force. These data should be considered in a validation step for a (re)-calibration of the RA scheme. To be noted is that the current RA is performed for single products which is a main flaw regarding the overall protectiveness. The overall protectiveness of the current RA scheme may be considered accordingly.

2.3 Spatially Explicit IBM: ALMASS Application to Birds and Mammals

The ALMaSS application for the vole was presented as by Dr Stephanie Kramer-Schadt (see Annex 1).

Short description of ALMaSS

ALMaSS: **A**nimal, **L**andscape and **Ma**n **S**imulation **S**ystem.

ALMaSS 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 platform is made of the following sub-models:

- Animals: IBM for skylark, vole, brown hare, carabid beetle and Linyphiid spider (among other for uses in PPP)
- Landscape Model: Spatial resolution of 1 m^2
- Man: Farming Management Model (e.g. timing of crop rotation)

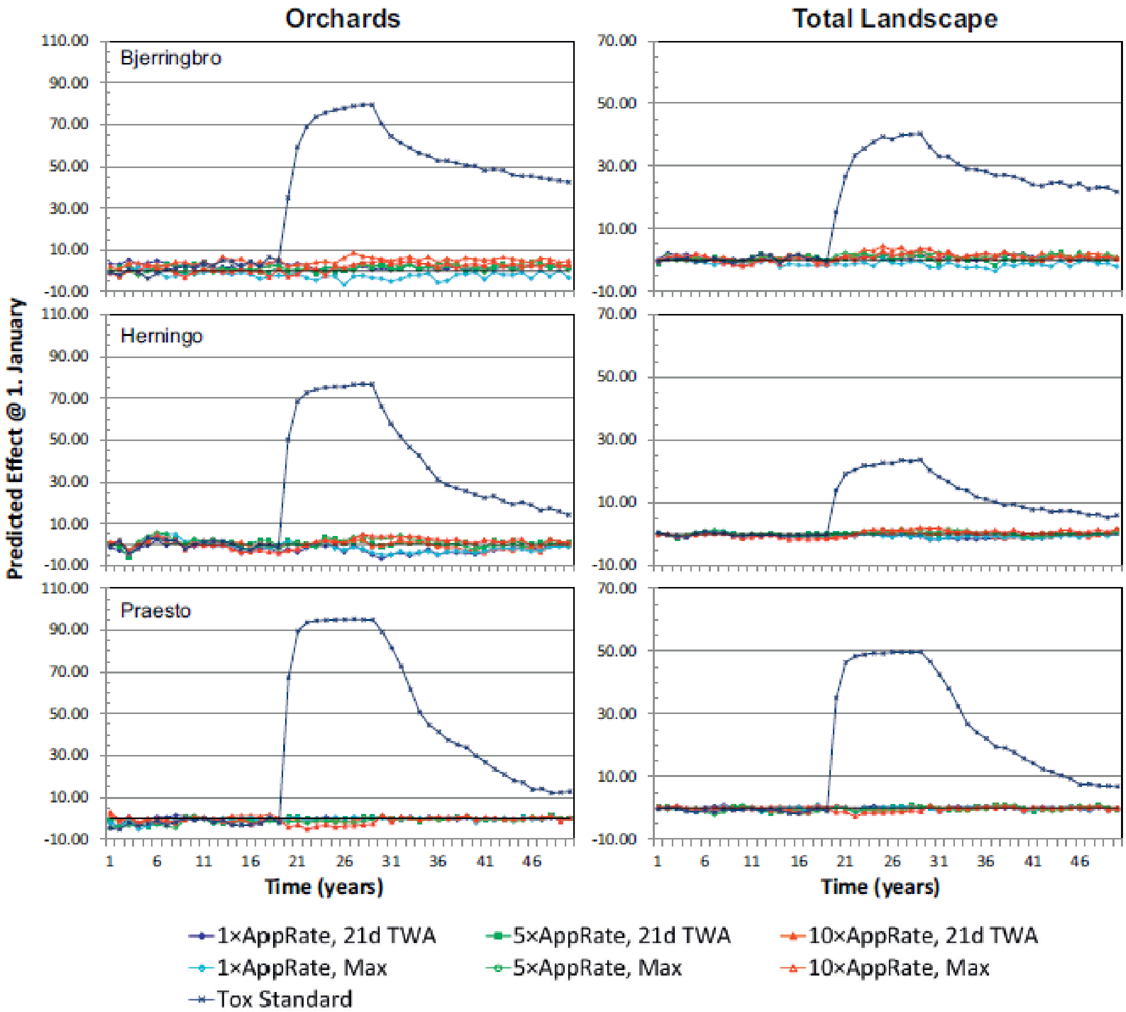
A brief application of ALMaSS for the vole was presented (see handouts of the talk).

Question directly linked to the case study

A participant noted that no effects were observed in the simulations for vole population dynamics over time (slide 22 of the presentation, see Fig. 53:) and suggested that this might be because the effects on population size were masked by the population dynamics (i.e. the toxic effects were smaller than the density dependent effects, thus no effect could be detected on the population size except for the toxic standard).

It was noted that also small fluctuations of population abundances may be of biological relevance.

Figure 53: Annex 2 – Application of ALMaSS to Field Voles



Predicted effects on field vole populations for three different landscape scenarios in orchards only (left side) and in total 10 x 10 km² landscape (right side). Effects on abundance from 10 years of pesticide treatment are shown. Tox-Standard is a positive control with a predefined mortality of 50 % for exposed voles. The figure was shown on slide 22 in the presentation of Stephanie Kramer-Schadt on ALMaSS at the symposium. Graph reproduced from Schmitt et al. (2015).

Validation

The question of the validation of ALMaSS was raised. Was the model only "tuned" (i.e. tested and adapted so that the simulation fits to the calibration data) or were the data independent? ALMaSS was not directly validated. This is maybe because long-term field studies with measurements on individuals are missing or in some cases even not feasible (e.g. carabids). Therefore, for most of the model applications (i.e. different "animal" sub-models) a rigorous validation is missing. In the case of the wood mouse, a pattern-oriented validation for simulations without pesticide exposure was conducted. In other cases, ALMaSS was "tuned" to a new situation.

For ALMaSS related to small mammals (e.g. sub-model on the vole), it was raised that information about rodenticide data could be used for validation.

Model implementation

Note

According to the EFSA Sci. Op. on GMP (2014), the model implementation is the conversion of the formal model into a computer model. In the modelling cycle, this also includes the verification of the code.

A modeler asked whether it is important that a model is open source.

It was confirmed that indeed this is an important criterion since many people (risk assessors and scientists) should be able to check the script / code of the models.

In addition, regulators raised the issue of a need for a "version control" to allow for rapid identification of a model version and of the changes made in the model as well as a preserved access to any previous version.

This was particularly relevant for models such as ALMaSS, which have a very long and complex code (i.e. thousands of lines), and in addition at least for ALMASS developed mainly by one person. However, in such a case, the fact that the model is "open source" would not be sufficient to support an accurate check. In the case of ALMaSS, the evaluation should focus on those parts of the code related to the critical / most relevant parts and parameters; the possibility of creating working groups of experts checking for those was mentioned.

Identification of Risk Mitigation Measures

A risk assessor asked whether models could be useful for the identification of Risk Mitigation Measures (RMM) (e.g. to define a protection zone) or for farm management.

The issue of focal species was raised. Currently the risk assessment is calculating a risk for a focal species, which should be representative for all other species. However, it might have a specific behaviour in the field and thus might not be representative for all other species.

A participant pointed out that in the current risk assessment, we only need to get the conclusion that there is no risk on one focal species, so this should apply for models as well.
It was acknowledged that landscape models could be used to identify or to validate the effect of buffer zones, i.e. to show higher and lower ecological risk with different RMM. The idea of using landscape-scale models for "bridging" risk assessment and risk management was rather well accepted in the meeting by the different participants.

However, it was strongly stressed that this might be considered possible only on the basis of validated models that can be trusted. The issues identified regarding the validation and the complex implementation (i.e. code development and verification) of ALMaSS were reiterated (see section above on implementation).

Validation of models, follow-up discussion

Since the ALMaSS model was only validated in parts, a risk assessor pointed out that this model may be of good quality but cannot be considered yet fit for the purpose of ERA. Then, a parallel to "global warming modelling" was made since this field of research is mainly based on models; but are these models validated? How did they reach such an acceptance by the community?

It was answered that in opposite to ALMaSS, which is almost the sole landscape-based platform discussed at the moment in the field of the PPP regulation, many models are available for climate change. So, the fact that many different independent models lead to the same predictions strongly enhanced the trust of the models in general.

Another participant indicated that in climate research, complex models can be checked by academic specialists because the research community is larger in this field than in Ecological Risk Assessment (ERA) modelling.

It was also stated that climate researchers cross-check their models. Models used in global warming modelling are in an “evaluation loop". They are running continuously. Everyday millions of measurements from the weather predictions are available to fit in the models. Thus, climate modelers are able to improve their models despite the fact that only a single field is existing (one earth!). In ERA in general there is the possibility to test / validate models against a variety of field situations. Accordingly, ERA is in a much better situation than climate change research in relation to testing of model variation.

It was also stressed by few participants that models used for climate change research are backed-up with reality whereas in ERA, predictions of models are only compared with results of experiments conducted in mesocosms (at least for aquatic organisms), which is a big difference. Models in ERA should thus be also backed-up with reality (i.e. monitoring data needed).

Post- Symposium Notes received during Commenting Phase

Some experiments / studies can provide good data for calibrating (or validating) models and this can also be considered as a type of “reality”, i.e. a real experimental system. Indeed, if suitable monitoring studies are available, they can contribute data to several parts in the modelling cycle (including validation). However, as monitoring data usually result from uncontrolled systems, the effort to measure all relevant parameters is high and the observed situations would then cover only single cases that not necessarily fit with the scope of the model.

A participant of the consortium pointed out that the database for global warming modelling differs from the one available for population models in ERA since in ERA there is a lack of field data. Field data from past situations for ERA do not exist to the same extent, especially for the time prior to the use of pesticides (no data are available), which might have been the most relevant when referring to baseline data. Indeed, first population field records are only from the 1950’s and concern only few species.

Exploring the intended uses / GAP

A risk assessor pointed out that in practice, they do not receive a range of answers or scenarios from the field studies and models presented in dossiers; actually, these show (in most cases) no effects. For better decision-making, risk assessors would need field studies or at least model calculations that predict a range of effects. Especially where field studies are not the best option (e.g. ethical issues, feasibility), models might help extrapolating from the lab to the field.

A modeler agreed with the necessity to identify clear effects with a model, as discussed for field studies, in order to increase “credibility” of the modelling approach; indeed, it can appear suspicious when only model calculations showing no effects are delivered. It was suggested that, if not provided, risk assessors should ask applicants to submit models run also with parameters showing clear effects (e.g. by simulating smaller application intervals or higher application rates which is similar to applying the margin of safety concept as for GUTS (see EFSA Sci. Op. on TKTD Modelling (2018, e.g. LPx/ EPX) and EFSA Sci. Op. on GMP (2014), page 67).

A risk assessor answered that usually applicants are not willing to submit extra data. Applicants have only to “defend” their GAP and therefore are not interested in spending time and money to explore additional environmental scenarios. In general, risk assessors need to have a solid justification to ask the applicant to deliver some extra information; to be noted: some consultants / applicants however are willing to deliver information and provide support to authorities.

It was suggested that authorities could be delivered the model in a software / format that is easy to use so that they can explore outputs under additional (worst-case) intended uses, i.e. get confidence that the model "works" and can predict various ranges of effects under various conditions. As suggested by another modeller, such an approach (i.e. using the model with various exposure modification factors multiplied with application rates until a risk is identified = the MoS concept) would also be necessary to assess the robustness of a model. Exploring additional (worst-case) environmental scenarios would also help to identify the "risk envelope", that would be useful to define the frame / focus of monitoring.

The benefit for authorities to explore themselves additional (worst-case) GAPs (assuming that the models would be delivered as a "software" easy to use) was in principle agreed, but it was also pointed out that currently most of the MS are lacking capacities and experts to manage this task. Therefore, some of these suggestions could be made mandatory for the applicant.

Post- Symposium Notes received during Commenting Phase

The models must be available in a form that they can be used also by the risk assessors – as it is the case for FOCUS models. However, as the risk assessors might not have the time to explore many additional scenarios, recommendations for specific ecological scenarios (or what kind of conditions should be covered) to be included in the submission are wished. For some products used in different crops, there are already many exposure scenarios; thus it would be feasible to restrict the selection to some of those already established exposure scenarios.

2.4 Spatially Explicit IBM: e-Vole Model Application

The eVole model application was presented as an example by Jeremias Becker (see Annex 1).

Model Outputs and Protection Goals

Justification for selection of case study: A modeler stressed that the e-Vole application presented is not a very recent application (before the EFSA Sci. Op. on GMP) and since then experience was gained to allow improvements for future applications. Although this is correct, it should be noted that at the start of the UBA project (2015), this study-case was recent (ZV1 DE zRMS).

In the case study presented, the model indicated that the vole populations could completely recover within 1 year after stopping the application of the fungicide. Indeed, after pesticide application, population density decreased to ca. 85 % of reference density for ca. 40 d, but less than 5% reduction in population density was observed by the end of the breeding season.

One of the concerns on the model application identified by the consortium was that the translation of the protection goals (PG) for mammals ("no visible mortality and no long-term repercussions" and "any mortality or reproductive effects unlikely" (EFSA 2009, EFSA PPR 2010)) into SPG is not well defined; thus it is difficult to use the quantitative model output in risk assessment by relating it to SPG (see slide 23). This led to a question from a risk assessor regarding the PGs, especially regarding the consideration of the recovery in such modelling approach. It was argued that for vertebrates in general (not specifically for voles), no long-lasting effects on populations are tolerable because during the season additional stressors and pesticides (other than the single one modelled) are present in the field. The vertebrate populations may be able to recover from the exposure to the specific substance addressed in the model, but they will be exposed to other substances as well. Please note that this aspect is i) not restricted to modelling but also applies to experimental approaches (e.g. field studies), and ii) possibly of higher relevance for vertebrates other than voles

Post- Symposium Notes

Effects on voles after exposure to pesticides other than rodenticides are mostly not shown in field studies.

Ecological modelling is generally proposed in dossiers as an additional line of evidence when other refinement options (i.e. residues decline, focal species, PT, PD…) are not sufficient to clearly exclude unacceptable effects. Therefore, risk assessors must make sure that the modelling option is actually reflecting the reality.

It was mentioned that if no effects should be tolerated, then models should not be applied to such a case study. A participant from academia replied that in this context recovery should indeed not be an option, but also pointed out that the predicted decrease of 10% by the model was also observed in the test; thus, it might be that these effects are reflected by the model.

Assessing uncertainties

In order to better identify how the toxicant does impact the densities of a population over time, a modeller suggested that margins of safety in the model predictions might be considered in population modelling by implementing the concept of multiplication factors such as the LPx approach proposed in the EFSA Sci. Op. on TKTD Modelling[2].

A modeller referred to slide 13 of the talk on eVole model, which is a comparison of simulations of the "basic model" of the vole (Wang, 2013) to observations from field studies. In this case, the predictions of the model were presented without any confidence interval. The consortium was asked how they coped with this issue while evaluating the models, especially regarding their validity.

This issue was acknowledged by the consortium. There was no concrete proposal but it was suggested to request further guidance and criteria for the reporting of the results as well as for the evaluation of the model (than given in the EFSA Sci. Op. on GMP).

Identification of RMM

A participant asked about the potential of eVole to identify relevant RMM. It was answered that because the landscape used in the model is quite simplified (field, field hedges and a refuge zone such as pasture), this model might not be best appropriate for this task. ALMaSS might be a better option. However, introduction of additional landscape complexity may come at the cost of reduced generalizability which is important when setting RMM for the GAP However, introduction of additional landscape complexity may come at the cost of reduced generalizability which is important when setting RMM for the GAP.

It was also said that the identification of RMM was not the initial aim of eVole. Instead, during the development much of the efforts were invested in the ecology of the species (e.g. distinction between adult and juvenile processes). Therefore, the model aims to better understand underlying processes of density effects observed at the population level.

Validation of population models

A risk assessor stated that the validation of modelled long-term population effects is (in principle) not possible and asked how to deal with this this issue.

Indeed, even under the hypothesis that a field experiment /monitoring would be performed to validate a population model, one can never be sure e.g. about the exposure of individuals to the pesticide since animals are moving. There is a dilemma since models could be useful to explore or simulate effects than cannot / should not be tested experimentally in the field; thus, these models cannot be fully validated. In other words, what cannot be measured, cannot be validated.

[2] LPx approach corresponds to a multiplication factor of the exposure profile that is necessary to reach a certain level of effect, e.g. 10% or 50%, i.e. LPx/EPx

A member from the consortium proposed to make use of historical data (even not necessarily with pesticides) for the validation of effect prediction in population modelling, if the required data cannot be generated otherwise. There was some agreement on this idea since in a model, e. g. the population recovery from a given acute effect (decrease in population size) is independent from the mode of action (toxicity module) that imposed this effect.

Further discussions are needed on how a model could be validated to be applicable for the risk assessment.

Post- Symposium Notes

The arguments stated above illustrate that a full validation of every single process or module in a population model is not feasible with (semi-) field studies. However, a population model can and should be subjected to a modular validation approach that follows the idea of pattern-oriented modelling: The outcome of each module can be tested under artificial but well-defined (laboratory) conditions to assess the structural integrity of a model. To test additionally the outcome of the full model with an available (semi-) field study, it is not necessary to know details such as the individual exposure of moving organisms in this study. Instead, the model is parameterized to the study conditions in the same way as it is done for a scenario to be used in ERA, and model predictions on the desired endpoint for risk assessment (e. g. change in population size over time relative to control runs) are compared to the observations from the study. Indeed, the more complex a modelled system, the larger the uncertainties in model predictions will be. Nevertheless, simulated confidence intervals should match those from (semi-) field studies, and this can be tested. If it is unethical to generate replicated observational data for the calculation of confidence intervals, an available data set may be split (bootstrapping) or historical data may be used (see above). If confidence intervals (CIs) are large but the lower ends of simulated CIs do not exceed those of observed CIs, the lower end of simulated CIs may be used in ERA.

Another main concern is that – by contrast with field studies – currently the behaviour of the animals is usually not considered in population modelling, although it would be of most relevance for important ecological processes (e.g. call of amphibians for mating).

Post- Symposium Notes

Many population models simulate the behaviour of individuals, from which the population dynamics emerge. However, most population models do not simulate pesticide-related changes in the individual's behaviour, typically because they are not known. In (semi-)field studies, population effects can be observed without knowing individual-level effects. Simulation studies may miss population effects due to the lack of knowledge on individual-level effects from which population effects may emerge.

A modeller from academia agreed that in principle it is hard to get the validation of the toxicant effects at the landscape level. However, it was suggested that for such spatial explicit population models the sensitivity analysis could consider/ focus on elements related to the landscape. For instance, simulations should not be restricted to one landscape. It might be relevant to manipulate the landscape to learn more about the model, e.g. to assess how the model reacts when the landscape structure, the available food resources or shelter are modified. If no change, then it might be more dependent from the toxicant.

A risk assessor acknowledged that this proposal might be relevant for the landscape structure. Then, usually the input parameters of the animals will remain the same but those related to the chemical (exposure and effects) can change. However, it was stressed that according to our knowledge, the way the effects of the toxicants are currently implemented in population modelling (i.e. toxicological submodel) is quite weak. Moreover, the higher the complexity of the model is, the more this Tox-part might get "lost". The validation of the part of the model which integrates the toxicological part and effects of other environmental factors on the population in the field is an issue.

A modeller suggested to conduct a step-wise validation, i.e. in first steps the model could be validated on a smaller scale and with ecologically simple species. For instance, the model could be developed for Daphnia and then extended to other long-lived species etc.

There was a common agreement that it is almost impossible to validate a model (in terms of correspondence of observed and predicted abundance over time) at the community level and in a complex landscape. Only the validation of some sub-models might be feasible. The validation of a model at field scale is probably not possible.

The consortium was asked about how they evaluated the population models in general and their validation in particular. It was answered that the evaluation was mainly based on the available information and recommendations provided in the EFSA Sci. Op. on GMP (2014),particularly the evaluation sheets and tables. About the validation, since it was missing in most case-studies used in the project, it could not be really evaluated.

Post- Symposium Notes

Simplified landscapes can be useful when tailored to a representative worst case that can be more easily analysed and communicated. With the landscape editor implemented in eVole, it is possible to create also more complex landscapes than in the example discussed at the workshop. However, due to its higher complexity, ALMaSS allows to simulate interactions between the population and the landscape in more detail, if this is deemed necessary output.

Ecological and landscape scenarios: Guidance for deriving such scenarios is wished.

Increase of realism and decision making

The proposals and statements on the validation were acknowledged by some risk assessors. However, as risk assessors need to trust the model, an appropriate validation is in principle needed. Indeed, as already mentioned at the beginning of the discussion, ecological modelling is a higher tier refinement option available in the Tiered approach.

Risk assessors have to assess whether the models are delivering sufficiently reliable results that will allow sufficiently safe decisions, i.e. need to avoid false negative is especially important in case of models showing no effects/ safe use. The major potential benefit from modelling is the increased realism of the predictions. However, some risk assessors doubted that the currently used spatially-explicit population models are actually able to increase realism due to a potentially unbalanced selection of processes to be considered e.g. on one hand, the spatial scale may enable to consider potential recovery but on the other hand, no interspecific interactions and limited additional stressors that may impair the potential recovery are included.

Some risk assessors question the usefulness / gain from a modelling approach compared to the current RA and are of the opinion that the use of ecological modelling should be limited to the "damage limitations" (i.e. RMM).

Other risk assessors are of the opinion that the main point is not that models could replace the current RA but to discuss what models could add to it.

A participant raised the challenge to balance between food production, landscape safety and preservation of biodiversity. In this view, landscape based ecological modelling might be a needed approach helping to make informed decisions, i.e. to inform on how the landscape should be better managed / organised?

A risk assessor commented that this is not only the task of risk assessors. Risks assessors have to correctly describe the risk and calculate whether effects are acceptable or not. They acknowledge that the risk might differ in different situations (i.e. agricultural landscapes). However, at the moment decision is taken based on the assumption that the scenario is realistic worst-case (i.e. then all scenarios are covered).

Facing the current biodiversity loss, it was pointed out that the decisions taken currently and the ERA framework may be wrong (e.g. focusing on authorising single pesticides independently of the agricultural contexts).

It was briefly discussed that although effects of pesticides in the field have been shown (EpiF report, 2005), the application of pesticides may not be the only factor responsible for this trend of biodiversity loss and that other anthropogenic factors - also acting simultaneously with pesticides - are also involved.
As a final point, some participants were wondering if it is worth to invest further effort for improving ecological modelling for the RA of PPP. Indeed, they were of the opinion that landscape based ecological modelling still needs much improvements and work. It might be a better strategy e. g. to invest efforts and resources in innovations of new pesticides, to further explore management options, to restructure agricultural landscape. Moreover, under climate change conditions achieving precise predictions may be tricky.

3 Wrap Up, Overall Discussion and Outlook

Strengths and limitations for GUTS, Debtox and Population models written below are wrap-ups of the discussions from Sessions 1 and 2 (first day), while questions 1 to 10 were discussed on the second day.

3.1. GUTS Models

What are the strengths and limitations of the GUTS models?	
Strengths Already assessed by EFSA PPR (2018) Enables to include all toxicity data (std. and non-std. data, all data from a dose/response relationship instead of only point estimates) -> potentially better link of exposure to effects Identifies problems with chemical (e.g. irreversible effects) Future-proof (e.g. mixtures, comparing species/chemicals)	**Limitations** - Every model is a simplification of the reality - GUTS address only survival and does not account for mortality due to long-term sublethal effects - No growth considered (but as no dilution factor included, suitable worst-case approach)

Question 1: Are GUTS models on survival fit for purpose and ready to use?	

Scientific aspects
- Mechanistic understanding instead of simple effect description
- Investigation of variable exposure in time and space in terrestrial environment (also applies to bird and mammal studies)
	Regulatory aspects
	- For aquatic organisms: ready to use according to EFSA Aquatic Guidance Document (EFSA PPR 2013).
	- For terrestrial organisms:
	  o effect part: not ready to use, as relevant endpoints are usually chronic and sublethal.
	  o exposure part: the application of GUTS (and also of classical dose-response fitting) is hampered as the exposure is often not well quantified, e.g. decay of substance not well captured in equation. In addition, the species is moving from more to less contaminated areas, i.e. physical terrestrial environment is patchier than in aquatic. This is also relevant for birds/mammals.
	- The “added value” of using TKTD compared to dose-response models, in terms of workload and improvement of ERA was questioned by some participants.
	- The power of prediction is not fully known since it is not the purpose of the GUTS application to simulate a community in the field (and thus no validation to the field).
	- GUTS is often proposed as Tier 2 using refined exposure profiles that present a number of uncertainties; but it is not proposed for Tier 1 under standard / constant exposure, whereas it could be useful, e.g. to detect irreversible effects.

3.2. DEBtox Models

What are the strengths and limitations of the DEBTox models?	
Strengths - Energy budget theory implemented for sublethal effects - Simplified models might also be useful - Add-my-pet database available for many species (full DEBtox) - Can be implemented in IBM as effect module for sublethal effects - Enables to combine all data/ different datasets (e.g. all data can be considered instead of only point estimates, datasets corresponding to different food levels, different chemicals) → adds to the information available for RA.	**Limitations** - EFSA PPR (2018): “not ready for use” (lack of relevant case studies/validation, lack of user-friendly software) - Many parameters to be calibrated (expertise required)

Question 2: Are individual model on sub-lethal effects (e.g. DEBtox) fit for purpose and ready to use?	
Scientific aspects - Understanding the mechanisms of action of different chemicals - The energy budget theory for sublethal effects is well established - Can be used for different species	**Regulatory aspects** - Could help linking (variable/field/monitoring) exposure data to sublethal effects - Agreement on environmental scenario is missing (. for scenarios other than standard tests settings, as defined in e.g. OECD test guidelines) - Lack of relevant case studies - Lack of user-friendly software (controlled version, stand-alone software); this would save time, resources and expertise. Type of software: R code or “press-button“ (risk assessors to test model by changing inputs parameters)? - Expertise needed to interpret the outcome

3.3. Population Models

What are the strengths and limitations of IB modelling?	
Strengths - Model mechanisms transparent - Stochasticity (from individual variability) is included - Exploration of long-term/ multi-stressors impacts on species potentially possible - Might be used to identify the population relevance of small effects detected in laboratory (e.g. Tier 1 tests) and to provide outcomes that are closer to the protection goals - Can be used for comparison of different scenarios (e.g. compare effects of a given exposure under different environmental conditions)	**Limitations** - Validation of the complete model often not sufficient/possible, e.g. ecological data are in most cases not available; feedback loop of the modelling cycle to the field situation would be needed (e.g. effects and fate monitoring) but is missing. - Compared to a standard test, usually the increased realism on the exposure side is clear (i.e. decreased exposure, decreased risk), while the increased realism on the effect side is variable (e.g. depending on the conditions of intra-species interactions set in the scenario, the sensitivity of the species under consideration may be increased). Model could thus be “unbalanced”, i.e. integrate mostly aspects decreasing the risk and not aspects increasing it.

	- Experience showed that toxic module sometimes not properly implemented or oversimplified (e.g. based on dose response curve of short-term lethal test, ignoring longer-term sublethal effects whereas they can be of high relevance for population dynamics)Because population effects emerge from individual-level effects, simulations tend to underestimate the real risk if not all individual-level effects that are relevant for the model processes are known (which is typically the case for sublethal effects such as behaviour modifications). - Quality and availability of data in many cases not sufficient. - Expert knowledge needed.

Question 3: are Individual Based population Models (IBM) fit for purpose and ready to use?	
Scientific aspects	**Regulatory aspects**
- Emerging properties can be identified - Identified emerging properties can be used for validation in a pattern-oriented modelling approach	- Validation for risk assessment needed and often not sufficient (feedback loop from monitoring and field studies missing); Availability of reliable data on ecological/ environmental stress for implementation in model is a key issue (e.g. data on population dynamics under different environmental conditions); such data is partly available in open ecological literature. - Could help bridging between risk assessment and management and monitoring - Added value in ERA needs to be demonstrated - Sensitivity analysis should point at the most influential factors (e.g. density dependency) - Identification / separation of toxicant effects from population natural variability is important (e.g. by comparing the same simulations settings with (treatment) and without (control) toxicant) - Conditions leading to recovery as emerging properties is not always clear, e.g. assessment of uncertainties around input parameters, selected processes - Potential for “biased” risk assessment approaches at higher tier, i.e. a robust risk assessment requires to consider mechanisms that increase the risk (e.g. indirect effects, sublethal effects, combined effects) and not only mechanisms that decrease the risk (e.g. toxicity exerted only through short-term lethal test, ignoring longer-term sublethal effects - No real pre-treatment/ historical data to establish baseline of the ERA and check for pesticide effects; this is not specific to modelling. - If the ecological model is robust, running the model with no exposure / effect (control conditions) would allow to produce data that simulates the situation pre-treatment. - Agreed environmental scenarios needed, e.g. with different set of conditions and levels of stress due to ecological, agronomical and ecotoxicological conditions (e.g. nutrient or food availability/vegetation cover/habitat quality) - Criteria for evaluation and presentation of results needed

3.4. Community Models

Not discussed

3.5. General Issues

Question 5: Validation of models is of utmost importance. How do you think this can be achieved?

- A full validation is possible in some cases (depending on species and groups). In other cases, some "modules" might be validated (e.g. ecological model), others not
- How the model is validated (i.e. with independent data) needs to be specified: which tier is used for validation? E.g., population model validated using data from "reference tier" (i.e. field) or "surrogate reference tier" (e.g. mesocosms)? Which specific applications?
- A proper validation of a population model designed to simulate effects in the field should ideally be towards the field; however semi-field data may be in some cases satisfactory.
- Data reported in ecological literature can offer possibilities to validate the ecological part of population models

Some issues

- Full validation in terms of effect studies in the field is sometimes not ethical (e.g. amphibians: protected species, whereas for this reason good models would be needed) or not possible (no field studies for terrestrial plants but models to fill up this gap!)
- Appropriate tox-modules considering relevant effects are needed (e.g. IBM can have complex exposure component but mismatch with effect side, e.g. change in behaviour as response to toxicant rarely considered) -> way forward: in models implement all relevant type of effects (e.g. behavioural changes) with all effect levels (e.g. not only NOAEL) observed in studies
- How to build confidence in model without classical validation?
- Pattern-oriented modelling could be used to validate the ecological model parts with independent data
- Explore the use of historical data / monitoring data, e.g. trace back population decline of species, assess available field studies
- Intensify monitoring activities for the feedback to ERA
- Calibration within the tiered RA and validation of the current (experiment –based) tiered approach with field data (considering a feed-back loop) is needed (issue of baseline: already disturbed system); but this issue is independent from the modelling (see also question 8)
- More trust in models validated with few field studies /monitoring studies. Some field studies (only available for industry as not submitted because they show risk) might be more useful for validation than studies showing no effect
- Endpoint in models to be translated into regulatory decisions, e.g. margins of safety

Question 6: How could population and community ecological models add to the risk assessment of pesticides?

- **Understand mechanisms?** Explore differences between lower and higher tier, combine different data addressing different aspects (e.g. short-term and longer-term toxicity, ecology of species, landscape, agricultural management) and from different origins (e.g. lab data, model data historical data)
- **Explore risk ranges?** Weight of evidence approach, exploration of different scenarios, time of onset of effects, margin of safety (models could show where risk starts). Address multiple exposure by different chemicals.
- **Demonstration of intended uses with acceptable risks /unacceptable risks?** Models should be used in a different way than experimental ("real") data (e.g. semi-field and field studies); e.g. not to show that a certain GAP induces no effects but rather for exploration of risks for different scenarios; critical for risk assessors to accept models since they are often delivered to demonstrate acceptable risk where conventional (current) approaches indicate unacceptable risks.
- Please note that models – as other approaches in ERA- should be linked to the SPG.

- → Main outcome: Models should not be used to replace field studies nor to replace the conventional approaches in RA. Considering both types of outcomes (models and experiments) in a weight of evidence approach seems most appropriate in decision making since both approaches address different types of uncertainties.

Question 7: Do you think models could be helpful to identify Risk Mitigation Measures?

- Spatially explicit IBM very useful to identify RMM, i.e. assess effectiveness of different RMM for different species in different landscapes
- This type of models already in use in the field of nature conservation (e.g. endangered species)
- Linking Risk assessment to risk management (NL Project). Effectiveness of RMM need to be identified already in RA
- MagPie Risk mitigation catalogue, no use of models to identify RMM at landscape scale (MagPie is largely dominated by emission reduction as mitigation)
- Can RMM be identified also with other models? Risk reduction factors needed (e.g. by 90%)?
- Link RMM needed for pesticide ERA to landscape management programmes under CAP.
- Identify “common” RMM useful for “greening” and pesticide ERA goals.

Question 8: Can you formulate specific model developments and requirements that would need to be addressed before possible implementation?

- Specific agreed models should be linked to specific questions in the risk assessment models
- Models could be used to assess/ explore the protectiveness of the current ERA (e.g. by exploring margins of safety when using different scenarios), but they should not fix the status quo/ define it
- Although this was not the purpose of the symposium, the need of calibrating the whole RA scheme with field (monitoring) data was raised as an important issue (also raised under question 5), especially in the context of the biodiversity decline. Ideally, this should be done before using model outcomes instead of conventional approaches in the RA. (**post symposium note**: models may help to better understand the reasons for biodiversity decline)
- Underlying factors are complex – but they need to be identified in order to be implemented in models (e.g. indirect effects as a cause for decrease of skylarks)
- GUTS: Exploration of the feasibility of validation criteria as formulated in the EFSA Sci. Op. on TKTD Modelling (2018).
- DEBtox: More case studies needed.
- (Spatially explicit) IBM: explore different scenarios, worst case is emerging from modelling, range of scenarios need to be agreed, explore ecological questions in relation to risk assessment outcome, more ecological data on species needed.
- Community models: (AQUATOX, huge calibration/validation efforts, used for retrospective RA, community models difficult to validate).

Question 9: How could model validation be supported and assessed at EU level?

- Support the set-up of ecological / chemical monitoring programs
- Validation on “field” or “mesocosm/semi-field” level? Check the model for the different levels of biological “complexity”
- It was questioned if accurate predictions are possible in very complex environments, if it is likely that complex models (with many parameters) might not react strongly to single parameter changes compared to less complex models; it was answered that the second item is part of the model’s sensitivity analysis

- The variability of responses from stochastic models should be accepted since it is also part of natural systems but how to report it and interpret it, by e.g. reporting range of responses?
- Retrospective identifications of chemical effects in multi-stressor environment to check if model identify the important factors
- Combining field studies/observations/post authorization monitoring/ event driven monitoring outcome with specific uncertainties addressed at "lower" level
- Explore ecological literature for data which allows to validate ecological parts of population models
- Pattern-oriented approach, combined output analysis
- Who is doing the validation and who is paying?

Question 10: How could model evaluation be agreed at EU level?

- Evaluation of models: should identify if the model includes all relevant processes that allow it to address a certain question (i.e. model domain's applicability), and if these processes are properly implemented and validated so that the model can be used
- Having a standardized set of models/modules would facilitate and help all involved parties, e.g. the evaluation of the "generic" (ecological) part of the model should be done before it reaches the desk of the single evaluator
- Evaluation of the "specific" part (ecotoxicological) related to the specific case/application? Active substance toxicity via EFSA conclusions /LoEP/DAR? Product, additional use evaluation, considering MoA?
- Evaluation of the outcome of the modelling exercise: should be done at MS, Zonal, Central zone levels?
- Need to establish an EFSA standing working group on model evaluation; however, priorities at EFSA are set by stakeholders via "self-tasking" or via EU COM
- Other EU groups (e.g. SETAC) could contribute to address the problems and add knowledge (e.g. help to identify evaluation criteria) and communicate to stakeholders; final decision on acceptance of criteria lies with EFSA

References

EFSA (2009): 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 (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): 92.

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.

EFSA PPR (2019): Technical report on the outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. Document No. EFSA supporting publication 2019:EN-1673.

EFSA Scientific Committee (2016): Scientific Opinion on recovery in environmental risk assessments at EFSA. EFSA Journal 14(2): 85.

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.

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.