

TEXTE

117/2025

Erarbeitung von Referenzwerten für Bodenorganismen für die Böden in Deutschland

Abschlussbericht

TEXTE 117/2025

Ressortforschungsplan des Bundesministerium für
Umwelt, Naturschutz und nukleare Sicherheit

Forschungskennzahl 3719 71 206 0
FB001136

Erarbeitung von Referenzwerten für Boden- organismen für die Böden in Deutschland

Abschlussbericht

von

Dr. Gabriel Salako, Dr. David Russell
Senckenberg Museum für Naturkunde, Görlitz

Andres Stucke, Dr. Einar Eberhardt
Bundesanstalt für Geowissenschaften und Rohstoffe (BGR),
Hannover

Im Auftrag des Umweltbundesamtes

Impressum

Herausgeber

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

Durchführung der Studie:

Bundesanstalt für Geowissenschaften und Rohstoffe
GeoZentrum Hannover
Stilleweg 2
30655 Hannover

Abschlussdatum:

August 2022

Redaktion:

Fachgebiet II 2.7 Bodenzustand, Bodenmonitoring
Dr. Frank Glante

Publikationen als pdf:

<http://www.umweltbundesamt.de/publikationen>

ISSN 1862-4804

Dessau-Roßlau, September 2025

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Kurzbeschreibung: Erarbeitung von Referenzwerten für Bodenorganismen für die Böden in Deutschland

Da in Deutschland keine flächendeckenden Referenzwerte für die Bodenbiodiversität insbesondere für Bodentiere vorliegt ist es schwierig einen sinnvollen Schutz für den Bereich Bodenbiodiversität zu entwickeln. Das Projekt Bodentierkartierung setzt an dieser Stelle an. Ziel des Projekts ist die Erarbeitung von großskaligen, räumlich expliziten Referenzwerten für die Bodenbiodiversität. Hierfür werden zoologische Daten der Bodentierfunddatenbank Edaphobase des Senckenberg Museums für Naturkunde Görlitz und Bodendaten der Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) genutzt.

Dem Projektplan folgend wurde die bodenbiologische Untersuchungslage der Bundesländer überprüft, um etwaige in der Edaphobase fehlende Fundortdatensätze aufzuspüren. Die Fundortdaten der Edaphobase wurden daraufhin durch Datensätze der Bodenübersichtskarte 1 : 200 000 der Bundesanstalt für Geowissenschaften und Rohstoffe ergänzt. Zur Vorbereitung der folgenden Modellierung wurden Daten zu Habitatemnung, Landnutzung, Klima und Bodeneigenschaften zusammengestellt.

Abstract: Development of reference values for soil organisms for the soils of Germany

Since there are no comprehensive reference data for soil biodiversity in Germany, especially for soil animals, it is challenging to develop a useful protection for soil biodiversity. Here, the project "Soil Animal Mapping" approaches the problem. The aim of the project is to develop large-scale, spatially explicit reference values for soil biodiversity. For this purpose, zoological data of the Edaphobase of the Senckenberg Museum of Natural History Görlitz and soil data of the Federal Institute for Geosciences and Natural Resources (BGR) are used.

Following the project agenda, the soil biology survey situation of the federal states was reviewed in order to detect any site data sets missing in the Edaphobase. The Edaphobase site data were then supplemented with data sets from the Soil Survey Map 1 : 200,000 of the Federal Institute for Geosciences and Natural Resources. Data on habitat suitability, land use, climate and soil properties were compiled in preparation for the following modeling.

Inhaltsverzeichnis

Abbildungsverzeichnis	8
Tabellenverzeichnis	10
Abkürzungsverzeichnis	11
Zusammenfassung	12
Summary	19
1 Einführung	25
1.1 Hintergrund	25
1.2 Stand der Forschung in Deutschland	30
1.3 Ziel	31
2 Material und Methoden	32
2.1 Räumlicher Umfang und Workflow	32
2.1.1 Zusammenstellung der Regenwurmdaten (Arbeitsschritt 1A)	32
2.1.1.1 Datenabfrage bei den Staatlichen Geologischen Diensten	32
2.1.1.2 Datenharmonisierung und -bereinigung	33
2.1.1.3 Parameter auf Gemeinschaftsebene (Gesamtabundanz und Artenreichtum)	35
2.1.2 Datensatz auf Artebene	37
2.1.3 Umweltdaten (Arbeitsschritt 1 B)	39
2.1.3.1 Zusammenstellung von Umweltdaten	39
2.1.3.2 Auswahl der ökologischen Variablen	40
2.1.4 Endgültige Variablenauswahl für die Modellkalibrierung (Arbeitsschritte 2 bis 4)	43
2.1.5 Modellkalibrierung und -anpassung (Arbeitsschritte 5 und 6)	46
2.1.5.1 Beschreibung der Modellalgorithmen	46
2.1.5.2 Modellanpassungen für GLM, GAM, GBM und RF	46
2.1.6 Prognosen (GIS-Kartenerstellung aus den Modellergebnissen) (Arbeitsschritt 7)	49
2.1.6.1 Visualisierung der Modellergebnisse	49
2.1.6.2 Diversitätsabschätzung	49
2.1.6.3 Gefährdungsabschätzung	49
2.1.7 Model assessment/evaluation (Arbeitsschritt 8)	50
3 Ergebnisse	52
3.1 Modellperformanz und endgültige Modellauswahl	52
3.2 Vorhergesagte räumliche Verteilung der Gesamtabundanz und des Artenreichtums der Regenwurmgesellschaften	57
3.3 Räumliche Verbreitung ausgewählter Arten	62

3.4	Größe des geografischen Verbreitungsgebiets der ausgewählten Regenwurmarten.....	66
3.5	Relativer Beitrag der Umweltvariablen zu den Modellvorhersagen und Verhalten von Gesellschaften und Arten gegenüber den Prädiktoren	68
3.5.1	Analysen auf Gemeinschaftsebene.....	68
3.5.1.1	Beiträge von Umweltprädiktorvariablen zu Modellergebnissen.....	68
3.5.1.2	Modellierter Response der Gesellschaften auf die Umweltprädiktorvariablen.....	68
3.5.2	Artspezifische Analysen	71
3.5.2.1	Beiträge von Umweltprädiktorvariablen zu Modellergebnissen.....	71
3.5.2.2	Modellierter Response der Arten auf Umweltprädiktorvariablen	73
4	Diskussion.....	83
4.1	Prognosen	83
4.2	Daten.....	88
5	Quellenverzeichnis	90
A	Anhang	99

Abbildungsverzeichnis

Abbildung 1:	Übersicht über den Arbeitsablauf der Studie.....	32
Abbildung 2:	Räumliche Verteilung der Regenwurmfunddaten.....	34
Abbildung 3:	Hauptkomponentenanalyse (PCA) der (standardisierten) quantitativen, d. h. numerischen vorausgewählten Prädiktorvariablen.	44
Abbildung 4:	Korrelations-Heatmap der 15 ursprünglich vorausgewählten standardisierten quantitativen Prädiktorvariablen.....	44
Abbildung 5:	Korrelationsmatrix der 13 standardisierten quantitativen Prädiktorvariablen nach Prädiktoreliminierung, basierend auf PCA- und Korrelationsanalyse	45
Abbildung 6:	Struktur und Entscheidungsfluss der IUCN-Kategorien zur Gefährdungsbeurteilung von Arten.....	50
Abbildung 7:	Initiale Gesamtabundanzvorhersage für Regenwurmgesellschaften durch drei unterschiedliche Modellalgorithmen.....	52
Abbildung 8:	Initiale Artenzahlvorhersage für Regenwurmgesellschaften durch drei unterschiedliche Modellalgorithmen.....	53
Abbildung 9:	Streudiagramm der vorhergesagten vs. beobachteten Gesamtabundanz von Regenwurmgesellschaften	53
Abbildung 10:	Streudiagramm der vorhergesagten vs. beobachteten Artenzahl von Regenwurmgesellschaften	54
Abbildung 11:	Boxplots der AUC-Performancewerte für die bewerteten Artenverbreitungsmodelle	55
Abbildung 12:	Boxplots der Cohen-Kappa-Performanzwerte für die evaluierten Artenverteilungsmodelle	55
Abbildung 13:	Boxplots der kombinierten AUC- und Cohens Kappa-Werte für die bewerteten Artenverteilungsmodelle	56
Abbildung 14:	Kombinierte AUC- und Cohen-Kappa-Wert-Verteilung der einzelnen Arten aller Modelle zusammen.....	57
Abbildung 15:	Gesamtabundanz der Regenwurmgesellschaften.....	59
Abbildung 16:	Artenreichtum von Regenwurmgemeinschaften.....	60
Abbildung 17:	Prognostizierte Erwartung zur Regenwurmdiversität.....	61
Abbildung 18:	Vorkommenswahrscheinlichkeiten von Arten weiter geographischer Verbreitung.....	63
Abbildung 19:	Wahrscheinlichkeit des Vorkommens von Arten mittlerer geographischer Verbreitung.....	64
Abbildung 20:	Wahrscheinlichkeit des Vorkommens von Arten mit eingeschränktem Verbreitungsgebiet und Beispiele für Arten mit besonderem Habitat oder endemische Arten	65
Abbildung 21:	Geographisches Verbreitungsgebiet von <i>A. caliginosa</i>	66
Abbildung 22:	Geographisches Verbreitungsgebiet von <i>D. attemsi</i>	67

Abbildung 23:	Beitrag (in %) der Umweltprädiktorvariablen zu Vorhersagen zur Gesamtabundanz und zum Artenreichtum der Regenwurmgemeinschaften	69
Abbildung 24:	Responsekurven (partielle Abhängigkeit) der vorhergesagten Gesamtabundanz von Regenwürmern auf die ausgewählten Umweltprädiktorvariablen	70
Abbildung 25:	Responsekurven (partielle Abhängigkeit) des vorhergesagten Artenreichtums von Regenwurmgesellschaften auf die ausgewählten Umweltprädiktorvariablen.....	71
Abbildung 26:	Relative Beiträge von Umweltprädiktorvariablen zu den modellierten Vorkommenswahrscheinlichkeiten ausgewählter Arten	72
Abbildung 27:	Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten der Arten auf die Variablen mittlerer Jahresniederschlag und Temperatur	74
Abbildung 28:	Kumulative modellierte Vorkommenswahrscheinlichkeiten in den Habitattypen, für die am häufigsten ein hohes Vorkommen von Regenwurmarten vorhergesagt wird	75
Abbildung 29:	Prognostizierte Habitatvorkommenswahrscheinlichkeiten der Arten über alle Habitatklassen.	76
Abbildung 30:	Übereinstimmung der prognostizierten Vorkommenswahrscheinlichkeit von <i>D. octaedra</i> (rechts) mit Habitattypen in Deutschland (links).	77
Abbildung 31:	Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten auf die Variablen Bodenfeuchte und pH-Wert.....	78
Abbildung 32:	Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten auf die Variablen organische Bodensubstanz und Gründigkeit des Bodens.....	79
Abbildung 33:	Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten („Auftretenswahrscheinlichkeit“) auf den Ton- und Schluffgehalt des Bodens als Proxy für die Bodentextur.....	81
Abbildung 34:	Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten („Vorkommenswahrscheinlichkeit“) von Arten auf die Bodenluftkapazität (Porosität) und die Lagerungsdichte des Bodens als Parameter der Bodenstruktur	82

Tabellenverzeichnis

Tabelle 1:	Für die Modellierung der potentiellen Artenverbreitung ausgewählte Regenwurmarten	38
Tabelle 2:	Liste vorausgewählter Umweltparameter, die als Prädiktorvariablen verwendet werden	41
Tabelle 3:	Europäische EUNIS-Klassifikationsebenen für Habitat-/Landnutzungstypen auf Ebene 1, die als Modellierungskategorien der Habitattypvariablen verwendet werden	42
Tabelle 4:	Liste ausgewählter Umweltparameter, die nach Autokorrelations- und Multikollinearitätstests als Prädiktorvariablen in die Modellalgorithmen eingehen	45
Tabelle 5:	Zusammenfassung der vier getesteten Algorithmen zur Modellierung der räumlichen Verbreitung von Regenwürmern in Deutschland	47
Tabelle 6:	Werte für Bestimmtheitsmaß und Übereinstimmungskoeffizient (C-Index) der unterschiedlichen Modellalgorithmen zur Vorhersage der räumlichen Verbreitung des Artenreichtums und der Gesamtabundanz der Regenwurmgesellschaften	52
Tabelle 7:	Mittlere AUC _{ROC} - und Cohens Kappa-Werte der bewerteten Modelle aller ausgewählten Arten	56
Tabelle 8:	Bewertung der Gefährdung der Arten basierend auf den IUCN-Kategorie-B-Kriterien 1 und 2 (NA: Funddaten für die Berechnung der Bereichsgrößen nicht ausreichend)	67
Tabelle 9:	Beitrag (in %) der Umweltprediktorvariablen zu Vorhersagen zur Gesamtabundanz und zum Artenreichtum der Regenwurmgemeinschaften	69
Tabelle 10:	Relative Beiträge von Umweltprediktorvariablen zu den modellierten Vorkommenswahrscheinlichkeiten von Arten ...	72
Tabelle 11:	Potenzielle Habitatpräferenzen der analysierten Arten (nach Prognose der Vorkommenswahrscheinlichkeitsmodelle)	77
Tabelle 12:	Datentabelle zu Abbildung 28: Vorkommenswahrscheinlichkeiten der Regenwurmarten in zehn Habitattypen	99
Tabelle 13:	Datentabelle zu Abbildung 29: Vorkommenswahrscheinlichkeiten der Regenwurmarten größer als der Schwellenwert 0,2 in zehn Habitattypen	99

Abkürzungsverzeichnis

AOO	Area of Occupancy (Areal einer Art)
BBodSchG	Bundesbodenschutzgesetz
BBodSchV	Bundesbodenschutzverordnung
BfN	Bundesamt für Naturschutz
BGR	Bundesanstalt für Geowissenschaften und Rohstoffe
BonaRes	Projekt: Boden als nachhaltige Ressource für die Bioökonomie
BÜK200	Bodenübersichtskarte von Deutschland im Maßstab 1 : 200 000
C	Kohlenstoff
C_{org}	Organischer Kohlenstoff
EFSA	European Food Safety Authority (Europäische Behörde für Lebensmittelsicherheit)
EOO	Extent of Occurrence
EUNIS	European Nature Information System
GBM	Generalized Boosting Model (= Boosted Regression Tree-Modell)
GLM	Generalized Linear Model
iDIV	Deutsches Zentrum für integrative Biodiversitätsforschung Halle-Jena-Leipzig
Ind. m⁻²	Individuen pro Quadratmeter (Maß der Abundanz)
KA5	Bodenkundliche Kartieranleitung (5. Auflage)
LUCAS	Land Use and Coverage Area frame Survey
MCP	Minimum Convex Polygons („konvexe Hüllen“)
N	Stickstoff
P	Phosphor
R	freie Programmiersprache für statistische Berechnungen und Grafiken
R²	Bestimmtheitsmaß
RF	Random Forest-Modell
RMSE	Root Mean Square Deviation, mittlere quadratische Abweichung (Standardabweichung)
SDM	Species Distribution Model (Artenverbreitungsmodell)
SOM	Soil organic matter
SOC	Soil organic carbon
Senckenberg Mu-seum	Senckenberg Museum für Naturkunde, Görlitz
UBA	Umweltbundesamt
WRB	World Reference Base for Soil Resources

Zusammenfassung

Die im Boden lebenden Organismen machen nicht nur einen großen Teil der terrestrischen Biodiversität aus, sondern tragen auch zu einem Großteil der Bodenprozesse und -funktionen bei, die die Ökosystemleistungen des Bodens ausmachen. Die enorme biologische Vielfalt im Boden und ihre funktionelle Bedeutung haben lassen die Notwendigkeit der Erhaltung der Bodenbiözonen erkennen. Viele nationale und EU-Rechtsvorschriften fordern ausdrücklich den Schutz und die Erhaltung der biologischen Vielfalt des Bodens und entsprechende Monitoringprogramme. Für ein Monitoring der Bodenbiodiversität kommen verschiedene Ansätze in Frage, die alle auf einer Bewertung der Ergebnisse anhand eines Vergleichs mit Referenzwerten basieren, die aus Felduntersuchungen an ausgewählten Referenzstandorten abgeleitet werden. Für die Erhaltung und den Schutz von Bodenlebensgemeinschaften sind jedoch evidenzbasierte Referenzwerte erforderlich, die auf einer breiteren räumlichen Skala die lokale und regionale Verteilung sowie die zugrundeliegenden Ursachen dieser Verteilung ableiten und Referenzwerte im Sinne von *standard operational ranges* für das Monitoring und die Bewertung der Bodenbiodiversität auf nationaler Ebene bereitstellen. Ein Ansatz zur Ableitung verallgemeinerungsfähiger Referenzwerte für die Bodenbiodiversität ist die Hochskalierung lokaler Beobachtungsdaten auf größere räumliche Maßstäbe unter Verwendung korrelativer Modellierungsmethoden.

In den letzten Jahren wurde mit der Entwicklung von Artenverteilungsmodellen (SDM) ein leistungsfähiges Instrument zur Bewertung und Vorhersage der räumlichen Verteilung der biologischen Vielfalt und ihrer potenziellen Einflussfaktoren entwickelt. SDMs modellieren statistisch die Korrelationen zwischen Arten oder Lebensgemeinschaften einerseits und Umweltparametern andererseits und nutzen diese Korrelationen, um das potenzielle Vorkommen der Arten bzw. Lebensgemeinschaften auf der Grundlage der räumlichen Ausprägung der Umweltparameter auf größere räumliche Maßstäbe hochzuskalieren (und zu kartieren) (Modelle der "Artenverteilung" oder der "Lebensraumeignung"). Dieser Ansatz bietet eine Lösung für die Entscheidungsunterstützung von Maßnahmen zur Erhaltung der biologischen Vielfalt. Mittlerweile gibt es verschiedene SDM-Modellierungsalgorithmen, wie z. B. *General Linear Models* (GLM), *Classification Tree Analysis* (CTA), *Multivariate Adaptive Regression Spline* (MARS), *Maximum Entropy* (MAXENT), *Random Forests* (RF) und *Generalised Boosted Regression Models* (GBM), die alle ihre Stärken und Schwächen haben. Zur Bewertung ihrer jeweiligen Fähigkeit Verteilung einer bestimmten Organismengruppe vorherzusagen, wird die Anpassungsgüte der Modelle u. a. anhand konventioneller Statistik wie R^2 , *Root Mean Square Error* (RMSE) sowie schwellenwertunabhängiger *Area under the Receiver Operating Characteristic Curve* (AUCROC) oder Kappa-Statistiken verglichen.

SDMs können auch zur Vorhersage der Verbreitungsgebiete von Arten verwendet werden. Während das "Verbreitungsgebiet einer Art" das Vorkommen eines Taxons innerhalb eines geografischen Gebiets beschreibt, ist die "Größe des geografischen Verbreitungsgebiets einer Art" (in km^2) die Fläche, auf der ein Taxon vorkommt. Die Größe des geografischen Verbreitungsgebiets wird als Indikator für den Bedrohungsstatus einer Art verwendet. Arten mit einem engen Verbreitungsgebiet sind viel stärker gefährdet als Arten mit einem größeren Verbreitungsgebiet. Zwei etablierte Maße für die Größe des geografischen Verbreitungsgebiets von Arten existieren: die (räumliche) Ausdehnung des Vorkommens (*extent of occurrence*, EOO) und das Verbreitungsgebiet (*area of occupancy*, AOO), die von der IUCN als Standardmaße für die Einstufung von Arten in Bedrohungskategorien verwendet werden.

Regenwürmer haben sich als äußerst wertvolle Indikatoren für die Bodenbiodiversität erwiesen, die zum Monitoring eines breiten Spektrums von Lebensräumen und Landnutzungen nützlich

sind. Obwohl die Rolle der Regenwürmer z. B. bei der Verbesserung der Bodenqualität seit langem bekannt ist, wurde bisher noch nicht untersucht, wie die Verbreitung dieser Bodentiergruppe entlang von Umweltgradienten beeinflusst wird und welche Faktoren wirksam sind. Regenwurmarten können in verschiedene ökologische Lebensformen eingeteilt werden (klassisch definiert als *epigäische*, die auf der Bodenoberfläche und in der Streu leben, *endogäische*, die im Mineralboden wühlen, und *anözische*, die in permanenten Röhren von der Bodenoberfläche bis in den Mineralboden hinein leben). Aufgrund ihrer unterschiedlichen Bindung an den Lebensraum Boden können sie tatsächlich unterschiedlich auf Umweltfaktoren reagieren. Soweit uns bekannt ist, wurde bisher in keiner Studie die Verteilung bestimmter Regenwurmarten mit mehreren Modellalgorithmen modelliert und räumlich kartiert, weder in Deutschland noch anderswo.

In der vorliegenden Studie wird die räumliche Verteilung von Regenwurmgemeinschaften (Gesamtdichte bzw. -abundanz, Artenreichtum) sowie die Wahrscheinlichkeit des Vorkommens und die Größe des geografischen Verbreitungsgebiets von zwölf ausgewählten Regenwurmarten in Deutschland vorhergesagt. Hierzu werden unter mehrere Modellalgorithmen geprüft und auf Grundlage der jeweiligen Modellleistung das geeignete Modell für die Projektion der Regenwurmverbreitung in ganz Deutschland ausgewählt. Die Studie folgt den ODMAP-Standardprotokollen für die Implementierung von Artenverbreitungsmodellen (SDMs) (Zurell *et al.*, 2020). Darüber hinaus werden mehrere Methoden (Experturteil, statistische Verfahren, ökologische Relevanzanalyse) zur Auswahl von Umweltprädiktoren verwendet, um eine Verzerrung bei der Variablenauswahl zu vermeiden und so ein breites Spektrum relevanter Umweltvariablen für eine robuste Modellierung der Verbreitung von Regenwurmarten abzudecken.

Algorithmen des maschinellen Lernens, wie *Random Forests (RF)*, *Generalized Boosted Regression Models (GBM)* oder die Maximum-Entropie-Methode (MAXENT) haben in jüngster Zeit bessere Ergebnisse erzielt als traditionelle Regressionsmodelle, wie *Generalized Linear Regression Models (GLM)* oder *Generalized Additive Regression Models (GAM)* (Elith *et al.*, 2006; Li & Wang 2013; Valavi, 2022). Obwohl RF bisher nur selten verwendet wurden und ihr Potenzial in SDMs nicht ausreichend genutzt wurde, hat ihre hohe Vorhersageleistung in letzter Zeit in angewandten ökologischen Studien Aufmerksamkeit erregt (z. B. Mi *et al.*, 2017). RF und GBM als Ensemble-Klassifikatoren bestehen aus mehreren alternativen Bäumen und verwenden diese bei der Entscheidungsfindung während der Erstellung von Modellvorhersagen (Li & Wang, 2013; Guisan *et al.*, 2017). Obwohl in dieser Studie die Leistung von GLM und GAM übertroffen wurde, sagt GBM nur innerhalb des dritten Quartils der Daten zur Gesamtabundanz und zum Artenreichtum voraus. Der Vergleich der Anpassungsgüte-Statistiken (R^2 , CI, AUC, Kappa), der beobachteten im Vergleich zu den vorhergesagten Datenanpassungen sowie der daraus resultierenden Karten der vorhergesagten Abundanz und des Artenreichtums durch alle Modelle veranschaulicht die gute Leistung von RF. So war dieser Algorithmus beispielsweise in der Lage, Vorhersagen über den Bereich des dritten Quartils der Dichtefelddaten hinaus zu treffen, einschließlich maximaler Dichten von über 600 Ind m^{-2} , und eine große Anzahl von Arten in den Daten zu erfassen. Frühere Studien bestätigen unsere Feststellung, dass Random-Forest-Algorithmen die beste Vorhersageleistung zeigen (z. B. Marmion *et al.*, 2008; Mi *et al.*, 2017; Valavi *et al.*, 2022). Es wurde jedoch eine Verbesserung der Anpassungsgüte für Dichtevorhersagen in RF sichtbar, nachdem zusätzliche Daten (aus Bayern) einbezogen wurden, was bestätigt, dass RF datenabhängig sein kann (s. a. Valavi *et al.*, 2021; Yiu, 2021). Die resultierenden partiellen Responsekurven, die die Beziehungen zwischen Gemeinschaften (oder Arten) und der Umwelt offenlegen, sind ein weiteres Beispiel dafür, wie RF-Modelle ökologisch relevante Ergebnisse liefern (Cutler *et al.*, 2007, Mi *et al.*, 2017).

Trotz der hohen Anpassungsgüte der RF-Modelle ist jedes Vorhersagemodell nur so gut wie die zugrunde liegenden Daten, die zur Kalibrierung verwendet werden. Mit über 20 000 Datensätzen von fast 1000 Fundorten können die biologischen Hintergrunddaten als umfangreich und ausreichend angesehen werden. Nur gut 40 % der Regenwurmdatensätze enthielten auch Daten zu den Umweltprädictoren, mit einer entsprechend hohen ökologischen Interpretierbarkeit der Funde. Die entsprechend großen Datenlücken wurden aufwändig durch externe Daten aus der Bodenübersichtskarte 1 : 200 000 von Deutschland (BGR 1997–2018) und weiteren Datenquellen ergänzt. Dies ist für einige Bodenparameter durchaus kritisch, für die flächenhafte Daten weiträumig auf Interpolationen beruhen, weil Böden kleinräumig sehr heterogen sein können und Interpolationen die Gefahr der räumlichen Überinterpretation der Umweltdaten mit sich bringen. Auch konnten nicht alle relevanten Parameter erfasst werden. So halten Creamer *et al.* (2019, in Baritz *et al.*, 2021) Indikatoren für die Qualität der organischen Bodensubstanz (z. B. C/N-, N/P-Verhältnis) auch für Bodenorganismen für sehr wichtig. Entsprechende Daten sind für Deutschland jedoch nicht flächendeckend verfügbar.

Es muss auch auf die potenziellen Schwierigkeiten bei der Verwendung externer Habitatdaten hingewiesen werden, da eine zeitliche Trennung zwischen Regenwurmbeobachtungen und Habitatübersichten eine auf Landnutzungsänderungen zurückgehende Fehlzuweisung enthalten kann. Glücklicherweise war der Habitattyp die häufigste Umweltmetadatengruppe, die mit den Regenwurmdaten erfasst wurde, so dass eine insgesamt befriedigende Modellkalibrierung gewährleistet war. Schließlich wurden nur abiotische Variablen als Prädiktorvariablen berücksichtigt; etwaige Wechselwirkungen mit anderen Organismen (d. h. zwischen Regenwurmarten, anderer Bodenfauna oder Mikroorganismen) wurden nicht berücksichtigt, da auch für andere Organismen(gruppen) keine entsprechenden Daten verfügbar sind (und Studien über mehrere Bodenorganismengruppen an denselben Standorten äußerst selten sind). Obwohl nicht alle potenziellen Einflussfaktoren auf die Verbreitung von Regenwürmern berücksichtigt werden konnten, enthielten die Modelle eine große Anzahl der wichtigsten Umweltparameter, von denen bekannt ist, dass sie die Fitness von Regenwürmern beeinflussen (z. B. Lee, 1985; Edwards & Arancom, 2022).

Obwohl die Modellvorhersagen noch nicht im Freiland validiert wurden (hier können die im Rahmen von Monitoringprogrammen des Bundes und der Länder vorgesehenen Erhebungen ein erster Schritt sein), bestätigen die veröffentlichten Zusammenstellungen der Autökologie der Regenwurmarten die Mehrzahl der Vorhersagen. Bemerkenswert ist der vorhergesagte Response der Arten auf den pH-Wert des Bodens, wobei viele acidophobe und einige acidophile oder -tolerante Arten mit einem Schwellenwert zwischen pH 4 und 5 ermittelt wurden. Graefe & Beylich (2003) berichten ebenfalls über eine solch starke artspezifische Differenzierung mit einem gemeinsamen Schwellenwert von pH 4,2, mit Ausnahme z. B. der acidophoben *A. longa* mit einem Schwellenwert von pH 5, der auch von unseren Modellen vorhergesagt wurde. Unsere Vorhersagen zu den Reaktionen der Arten auf die Bodenazidität werden auch weitgehend bestätigt, z. B. von Sims & Gerard (1999), Jänsch *et al.* (2013), Krück (2018) und Sherlock (2018). Diese Autoren sowie Römbke *et al.* (2018) und Lehmitz *et al.* (2016) beschreiben auch artspezifische Präferenzen für den Gehalt an organischer Bodensubstanz (SOM), die durch die Modellvorhersagen fast vollständig bestätigt werden. Einige dieser Autoren berücksichtigten auch Präferenzen für den Tongehalt, die im Allgemeinen, aber nicht immer, von den hier vorgelegten Modellvorhersagen bestätigt werden. So berichten Jänsch *et al.* (2013) über die Vorliebe von *D. octaedra* für Böden mit niedrigem Tongehalt und die leichte Vorliebe von *A. chlorotica* für Tonböden, die beide durch unsere Ergebnisse widerlegt werden. Auch die von uns vorhergesagte positive Reaktion von *L. terrestris* auf Böden mit geringerem Ton- und Schluffgehalt steht im Widerspruch zu der Einschätzung von Sims & Gerard (1999) und Sherlock (2018), dass diese

Art lehmreiche Böden bevorzugt (diese Autoren beziehen sich allerdings auf britische Populationen).

Die vorliegende Studie bestätigt die Aussagen früherer Arbeiten zu den Auswirkungen von Niederschlag und Bodenfeuchtigkeit; diese Variablen sind für die Zunahme der Population und die Verteilung erwachsener Regenwürmer nach Regenfällen verantwortlich, insbesondere in relativ trockenen Gebieten oder nach langen Dürreperioden (Lavelle, 1978; Lavelle & Spain, 2005; Kalu *et al.*, 2015; Rajwar *et al.*, 2022). Phillips *et al.* (2019) ermitteln mit einfacheren statistischen Methoden das Klima (insbesondere den mittleren Jahresniederschlag und die Temperatur) als fast ausschließlichen Einflussfaktor auf die Regenwurmgemeinschaften (Gesamtdichte, Artenreichtum) auf globaler Ebene. Unsere Studie bestätigt zwar die kombinierte Rolle von Temperatur, Niederschlag und Bodenfeuchtigkeit, identifiziert aber auch den Lebensraumtyp, den pH-Wert des Bodens und die organische Bodensubstanz als wichtige Einflussfaktoren, die zusammen etwa 70 % zu den Vorhersagen beitrugen. Da das Klima auf globaler Ebene z. B. auch die natürliche Vegetation (die zu Biomen und Lebensraumtypen führt) und teilweise auch die Bodenengehese beeinflusst, ist es plausibel, dass statistische Methoden das Klima gegenüber anderen Umweltparametern als treibende Kraft auf globaler Ebene identifizieren. Unsere Studie auf regionaler Ebene zeigt jedoch auch die Bedeutung von Lebensraum- und Bodenparametern als zusätzliche Faktoren für die Verbreitung von Regenwürmern. Im Gegensatz dazu bezweifeln Hodson *et al.* (2021), dass Bodenfaktoren ein wichtiger Faktor für die Regenwurmdichte und -verteilung sind, insbesondere auf nationaler Ebene. Die maßgeblichen Faktoren für die biologische Vielfalt der Regenwürmer sind also offensichtlich skalenabhängig; Klimaparameter sind auf globaler und regionaler Ebene wichtig, während Lebensraum- und Bodenfaktoren auf kleineren räumlichen Skalen an Bedeutung gewinnen. Auf lokaler Ebene (wo ähnliche klimatische Bedingungen herrschen) dürften die letztgenannten Faktoren am wichtigsten sein, und auch anthropogene Landnutzungsmaßnahmen werden die Artenvielfalt der Regenwürmer zunehmend beeinflussen.

Die vorliegende Studie sagt Vorkommenswahrscheinlichkeiten jenseits der traditionellen Nutzungstypen Wald, Grünland und Acker vorher, indem alle terrestrischen EUNIS-Level-1-Lebensraumtypen, einschließlich mariner (d. h. Inseln), küstennaher und städtischer Standorte, einbezogen wurden. Diese Vorhersagen stimmten in hohem Maße mit den vorgeschlagenen Klassifizierungen der Verbreitungsgebiete auf der Grundlage der Verbreitungsvorhersagen überein. So wurde für die meisten Arten mit großem Verbreitungsgebiet eine weite Verbreitung in vielen Regionen Deutschlands und ein Vorkommen in vielen verschiedenen Lebensraumtypen prognostiziert, oft mit Wahrscheinlichkeiten von mehr als 50 – 60 %, was auf ihren ökologisch generalistischen Charakter hinweist. In der Roten Liste Deutschlands werden alle diese Arten als sehr häufig aufgeführt (Lehmitz *et al.*, 2016), und in der Literatur werden sie als euryök bzw. wird ihr Vorkommen als ubiquitär beschrieben (z. B. Sims & Gerard, 1999; Jänsch *et al.*, 2013; Römbke *et al.*, 2018; Sherlock, 2018), was unsere Ergebnisse bestätigt. Obwohl *L. terrestris* im Allgemeinen als eurytop angesehen wird, wird gelegentlich auf eine leichte Präferenz für Grünlandstandorte hingewiesen (Sims & Gerard, 1999; Jänsch *et al.*, 2013; Sherlock, 2018), was durch unsere Vorhersagen bestätigt wird, die aber in sehr ähnlichem Maße auch Waldhabitatem vorhersagen. Die Art gilt als störungsintolerant (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018), was ihre geringe Wahrscheinlichkeit für natürliche Sonderstandorte (z. B. Auen, Moore) sowie anthropogen stark beeinflusste Lebensraumtypen erklären könnte. Vorsicht ist u. U. bei den Aussagen zu *A. caliginosa* geboten, die kryptische Arten enthalten kann (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018); dies ist jedoch eine Frage der Taxonomie und der Artbestimmung und nicht der Modellierungsverfahren.

Den von uns als Arten mit mittlerem Verbreitungsgebiet eingestuften Arten wird ebenfalls ein weites Vorkommen in Deutschland vorhergesagt, wenn auch oft mit deutlich geringerer Wahrscheinlichkeit. In der Roten Liste Deutschlands werden sie alle als *häufig* geführt. Obwohl für sie ein Vorkommen in vielen verschiedenen Lebensraumtypen prognostiziert wird (euryöke Arten), scheinen sie stärker habitatabhängig zu sein und ein Optimum in bestimmten Lebensraumtypen zu haben. So wurde beispielsweise für *A. chlorotica* ein stärkeres Vorkommen auf landwirtschaftlich genutzten Flächen (Acker oder Grünland) vorhergesagt, was auch aus Beobachtungsdaten hervorgeht (z. B. Jänsch *et al.*, 2013; Römbke *et al.*, 2018). Andererseits wird *D. octaedra* hauptsächlich in Waldhabitaten vorkommend vorhergesagt, was mit Erkenntnissen von z. B. Jänsch *et al.* (2013), Römbke *et al.* (2018) und Sherlock (2018) korrespondiert. In Anbetracht der acidophilen Natur der Art ist eine Präferenz für Nadelwälder wahrscheinlich (vgl. Sherlock, 2018). Während *A. castaneus* eher ein Generalist zu sein scheint, zeigen die vorliegenden Ergebnisse die höchsten Vorkommenswahrscheinlichkeiten in Wäldern und Auen, was u. a. Jänsch *et al.* (2013), Römbke *et al.* (2018) und Krück (2018) widerspricht, die eine Präferenz auch für Grasland sehen. Interessanterweise wurde *A. longa* von unseren Modellen auch als Generalist vorhergesagt, der in verschiedenen Lebensraumtypen vorkommt, aber in feuchteren Lebensräumen (z. B. Inseln, Küsten, Überschwemmungsgebieten, Mooren) fehlt. Dies wird von Krück (2018) bestätigt, der *A. longa* eine Vorliebe für trockenere Lebensräume bescheinigt, im Gegensatz zu Sims & Gerard (1999), die ein Vorkommen der Art in Überschwemmungsgebieten Großbritanniens feststellten.

Die Arten mit eingeschränktem Verbreitungsgebiet weisen alle eine auf bestimmte Regionen und Lebensraumtypen beschränkte Vorkommenswahrscheinlichkeit auf. In der Roten Liste Deutschlands sind sie alle als *selten* oder *sehr selten* aufgeführt. Zum Beispiel liegen die höchsten Verbreitungswahrscheinlichkeiten von *D. attemsi* eher in den Hügelländern und Mittelgebirgen; die höchsten Wahrscheinlichkeiten wurden für die Nutzungsart Acker (und in zweiter Linie für Wälder) vorhergesagt. Dies steht im Widerspruch zu früheren Arbeiten, mit Ausnahme von Sherlock (2018), der Vorkommen von *D. attemsi* in Wäldern Großbritanniens feststellte. *A. eiseni* wird von den Modellen als am wahrscheinlichsten in Wäldern (in Mittel- und Süddeutschland) vorkommend vorhergesagt, wie auch Römbke *et al.* (2018) und Lehmitz *et al.* (2016) berichten. *A. limicola* ist als hydrophil bekannt (Sims & Gerard 1999; Lehmitz *et al.* 2016; Krück 2018; Römbke *et al.*, 2018; Sherlock, 2018). Dementsprechend sagen die Modelle voraus, dass die Art mit hoher Wahrscheinlichkeit in Überschwemmungsgebieten sowie im Grasland und in Wäldern vorkommt – die sich möglicherweise ebenfalls in Überschwemmungsgebieten u. ä. befinden, deren Funddaten aber möglicherweise fälschlich allgemeineren Lebensraumtypen zugeordnet wurden. Die Modelle sagen voraus, dass *A. limicola* vor allem in Westdeutschland (und am stärksten im Rheintal) vorkommt, was zu den Befunden von Krück (2018) passt, dass die Art in Nordostdeutschland selten vorkommt. *L. badensis* ist eine in Deutschland endemische Art, die in den Wäldern des Hochschwarzwalds (Südwestdeutschland) vorkommt (Lehmitz *et al.*, 2016), was sich auch in den Modellvorhersagen zeigt.

Interessant sind die wenigen Arten, die nach den Modellierungen in Ausnahmehabitaten vorkommen. So wurde für *A. limicola*, *L. rubellus* und *L. castaneus* ein Vorkommen in marin beeinflussten Lebensräumen (d. h. auf Inseln) und für *A. caliginosa*, *A. chlorotica* und *A. limicola* in Küstengebieten vorhergesagt; alle jedoch mit einer geringen (<35 %) Vorkommenswahrscheinlichkeit. Das weist auf ein lückenhaftes Vorkommen in diesen Lebensräumen hin. Auffallend ist die große Anzahl von Arten, die mit mäßiger Wahrscheinlichkeit in städtischen, industriellen und anderen anthropogenen Gebieten vorkommen. Diese neuen Ergebnisse können bei der Bewertung von Erhebungen zur biologischen Vielfalt des Bodens in solchen Gebieten helfen.

Auf der Ebene der Lebensgemeinschaften ist die geografische Variation und die inverse Beziehung zwischen Artenreichtum und Gesamtabundanz, wie sie in dieser Studie (hauptsächlich für Nordostdeutschland) festzustellen ist, ein häufiges Phänomen in der Ökologie. Gesamtabundanz und Artenreichtum verhalten sich oft gegenläufig, so dass ein Gebiet eine hohe Individuendichte, aber einen geringen Artenreichtum aufweist (Verberk *et al.*, 2011). Eine Region mit ungünstigeren Umweltbedingungen ermöglicht ggf. nur das Vorkommen einiger weniger Arten, diese aber in durchaus großen Populationen (hohe Individuenzahlen), u. U. gerade aufgrund geringerer Konkurrenz durch andere Arten (Groves, 2022). Dies könnte möglicherweise die hohe Individuendichte, aber den geringen Artenreichtum erklären, der für Nordostdeutschland (insbesondere Brandenburg) vorhergesagt wird, wo vergleichsweise trockenere, sandige Böden vorherrschen und wo vorwiegend Nadelbaumforste stocken. In diesem Zusammenhang sind die hohen Vorkommenswahrscheinlichkeiten von *D. octaedra* und *L. rubellus* in diesen Gebieten auffällig, bei des epigäische, acidophile (oder -tolerante) Arten mit einer vermuteten Affinität zu Wäldern. Eine Erhebungen der biologischen Vielfalt im Boden haben oft hohe Populationsdichten von sehr wenigen epigäischen Arten in Wäldern auf sandigen Böden gezeigt. Andererseits werden für die bayerischen Alpen und das Rheintal als einige der wenigen Regionen mit hoher Regenwurm-Biodiversität (sowohl Gesamtabundanz als auch Artenreichtum) in Deutschland vorhergesagt. Das Rheintal ist für reiche Böden und eine hohe allgemeine Biodiversität bekannt, und die Vorhersagen für die alpinen Regionen gehen konform mit der Alpenkonvention (2014), die die Alpen aufgrund ihres natürlichen Zustands und der geringe(re)n menschlichen Störungen als eine der reichsten Regionen Europas in Bezug auf die Vielfalt von Pflanzen und Tieren anerkennt.

Die Größe des Verbreitungsgebiets ist seit langem als guter Indikator für die Bewertung der Bedrohung einer Art anerkannt. Der Bedrohungsstatus der Regenwürmer in Deutschland ist in der entsprechenden Roten Liste dokumentiert (Lehmitz *et al.*, 2016). Die Kartierung der räumlichen Verbreitung der Regenwurmarten und die Bestimmung der Flächengröße erlauben, die Arten in Verbreitungsgruppen (Arten mit großem, mittlerem und kleinem Verbreitungsgebiet) einzuteilen und Aussagen zum Bedrohungsstatus und Schutzbedarf der Arten abzuleiten (IUCN 2012 a, b; 2022).

Keine Art wurde nach den Kriterien des Verbreitungsgebiets als bedroht eingestuft (die Verbreitungsgebiete aller Arten überstiegen den Mindestschwellenwert von 20 000 km²), was darauf hindeutet, dass die untersuchten Regenwurmarten in Deutschland potenziell weit genug verbreitet sind, als dass sie durch Einzelereignisse aussterben (IUCN 2012, a, b). Dennoch zeigt ein Vergleich der prognostizierten Verbreitungskarten und des berechneten Verbreitungsgebiets (Area of Occupancy, AOO), dass bestimmte Arten aufgrund ihres eingeschränkten Verbreitungsgebiets zur Anlass zur Sorge geben, z. B. *L. castaneus*, *D. octaedra* und *A. longa* (als gefährdete Arten) sowie *A. eiseni*, *D. attemsi* und *A. limicola* (als stark gefährdete Arten nach den AOO-Kriterien) oder weil sie in Deutschland endemisch sind, wie *L. badensis*. Über den Status von *A. eiseni* in Deutschland existieren unterschiedliche Meinungen; während Bouche (1972) und Graff (1953) die Art in Frankreich bzw. Deutschland als selten einstufen, stufen Römbke *et al.* (2017) die die Art als häufig ein. Unsere Ergebnisse unterstützen teilweise die älteren Befunde von Graff (1953) und Bouche (1972) sowie die Zwischenposition von Lehmitz *et al.* (2016), die die Art als mäßig häufig einstufen (die vorliegende Studie bevorzugt den Begriff des eingeschränkten Verbreitungsgebiets).

Dagegen müssen die hier gemachten Vorhersagen über ein begrenztes Vorkommen von *A. eiseni* in hessischen Wäldern (Römbke *et al.*, 2017) und einigen anderen Gebieten zurückhaltend beurteilt werden. Die Art wird wahrscheinlich arboreal und kortikal, und die begrenzten Beobachtungsdaten (auf denen die Modelle basieren) könnten methodisch verzerrt sein, da die üblichen

Methoden zur Regenwurmextraktion die bevorzugten Mikrohabitatem dieser Art möglicherweise nicht ausreichend erfassen (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018). *A. limicola* ist die einzige hier untersuchte Art, die in der deutschen Roten Liste der Regenwürmer als gefährdet aufgeführt ist. Sein prognostiziertes Vorkommen in Feuchtgebieten, das seinen Status als hydrophile Art bestätigt, sowie seine prognostizierte begrenzte Verbreitung in Deutschland spiegeln seinen Gefährdungsstatus wider. Die übrigen Arten sollten bei künftigen Erhebungen zur biologischen Vielfalt im Boden als Schwerpunktarten betrachtet werden.

Besonders hervorzuheben ist *L. badensis*, eine endemische Art, die in der Hochschwarzwaldregion vorkommt und wahrscheinlich gefährdet ist (Lehmitz *et al.*, 2016). Obwohl keine ausreichenden Funddaten für die Berechnung von EOO oder AOO zur Verfügung standen, wurde für diese Art ein sehr eingeschränktes und enges Verbreitungsgebiet vorhergesagt, was ihrer Einstufigung als gefährdet untermauert. Diese Art verdeutlicht auch einen wichtigen Aspekt der Verbreitungsmodellierung: Obwohl die Modelle ein potenzielles Vorkommen z. B. in den bayerischen Alpen vorhersagen, wurde die Art dort noch nie nachgewiesen. Die Modellergebnisse zeigen demnach nur eine hohe potenzielle Lebensraumeignung für die Art in den Alpen, sind aber kein Beleg für ihr dortiges Vorkommen.

Summary

Soil-dwelling organisms comprise not only a major portion of terrestrial biodiversity, but also drive a majority of the soil processes and functions that lead to terrestrial ecosystem services. The immensity of soil biodiversity and its functional importance has led to the appreciation of the conservation needs of soil biocoenoses, whereby much national and EU legislation specifically calls for protecting and conserving soil biodiversity through explicit monitoring programs. Many approaches have been outlined for monitoring soil biodiversity, whereby all of these base an assessment of monitoring results on a comparison with “reference values” derived from field surveys in selected reference sites. However, conservation and protection of soil biotic communities require evidence-based baseline information – derived at broader spatial scales – on their local and regional distributions as well as the drivers of this distribution, which is necessary for formulating reference values (“standard operational ranges”) for national-level soil-biodiversity monitoring and assessment. An approach for deriving generalizable soil-biodiversity baselines is to upscale local observational data to broader spatial scales using correlative modelling methodologies.

In recent years, the development of species distribution models (SDMs) has advanced a powerful tool for assessing and predicting biodiversity distribution and its potential drivers. SDMs statistically model species’ or community’s correlations with environmental parameters and use these correlations to upscale (and map) their potential occurrences to larger spatial scales based on the spatial distribution of the environmental parameters (“species distribution” or “habitat” suitability” models). Such mapping methodologies represent a core solution for decision support of biodiversity conservation policies. Various SDM modelling algorithms exist today, such as Generalised Linear Models (GLM), Classification Tree Analysis (CTA), Multivariate Adaptive Regression Spline (MARS), Maximum Entropy (MAXENT), Random Forest (RF) and Generalised Boosted Regression Model (GBM), among others, each of which possess inherent strengths and weaknesses. To assess their specific capacity to accurately predict the distribution of a specific organism group, model goodness-of-fits are compared using conventional statistics such as R², root mean square error (RMSE) as well as threshold-independent “area under the receiver operating characteristic curve” (AUCROC) or Kappa statistics, among others.

SDMs can also be used to predict species’ distribution ranges. While a “species distribution range” describes the occurrence of a taxon within a geographic area, “species geographic range size” is the geographic area (in km²) in which a taxon is found. Geographic range sizes have been used as an indicator for the threat status of a species, narrow-range species being much more vulnerable than species with wider range sizes. Two established metrics measure species geographic range size: extent of occurrence (EOO) and area of occupancy (AOO), which have been adopted by the IUCN as standard practice for classifying species into threat categories.

Earthworms have been identified as being highly valuable indicators of soil biodiversity, useful in monitoring a broad range of habitats and land-uses. Although the role of earthworms in, e.g., improving soil quality has long been established, assessing how this soil-animal group’s distribution is influenced along environmental gradients and quantifying the drivers of such distribution has not yet been undertaken. Earthworm species can be classified into different ecological life-form types (classically defined as “epigeic” living on the soil surface and litter, “endogeic” burrowing in the mineral soil, and “anecic” living in permanent burrows between the soil surface and the mineral soil). Due to their different associations with the soil habitat, they may indeed react differently to important environmental drivers. To the best of our knowledge, no study has yet been conducted to model and spatially map the distribution of specific earthworm species testing multiple model algorithms, in Germany or elsewhere.

The current comprehensive study statistically modelled the potential earthworm distribution throughout Germany, evaluating both general community parameters (species richness and total density) as well as the occurrence probabilities of individual species. The general goal was to predict and map earthworm spatial distribution to order to determine their conservation needs and provide baseline reference data for Germany. It therefore sought to achieve the following objectives: (1) apply and compare correlative modelling techniques to predict and map the spatial distribution and geographic range of earthworm community parameters as well as carefully selected earthworm species in Germany, (2) identify and determine the importance of environmental predictors (drivers), and (3) evaluate whether earthworm life-form types are predicted to generally react differently to environmental drivers. The results are intended to provide background reference values for comparison with future monitoring results, as a first step in defining site-specific soil-biodiversity baseline values. As a final goal, this study also strove to (4) establish the species distribution pattern and density relationships to determine the conservation needs of earthworms in Germany.

SDMs are based on the correlation between a taxon's true observed distribution and the environmental conditions of the observed sites of occurrence. For this, observational data on earthworms in Germany was downloaded from the soil-biodiversity data warehouse "Edaphobase". Additional earthworm data was obtained from data providers or by consulting provider's reports and publications. This resulted in a dataset consisting of 22,134 individual data records (rows) from 992 locations (sites of occurrence) throughout Germany, which can be considered to be sufficiently large for distribution-modelling purposes. All available metadata concerning soil, habitat types, climate, etc. were downloaded with the earthworm data, being specifically linked to the earthworm data per specific site of occurrence. For community-level modelling, the data was aggregated to "total earthworm density" (harmonized to individuals per square meter [ind. m⁻²]) and "earthworm species richness" (average number of species found occurring in a site). For species-level data, data was transformed to "occurrence" (yes/no [= 1/0] for each species and site. The Edaphobase data included occurrence information on 45 valid species. While all were used to model community density and richness, only 12 species were selected for modelling species-level spatial distribution due to limited data on other species as well as available time and resources. Selection choice was based on (1) sufficient data for modelling purposes, (2) common species based on literature and expert knowledge, (3) species with apparent unique habitat preferences, (4) species considered or speculated to have limited distribution and (5) species considered to be endemic.

A number of environmental predictor variables known to be physiologically and ecologically importance for earthworms were initially selected. As many Edaphobase records did not contain all of these variables, external data was used to augment data gaps. Climate data was downloaded from Climatologies at High resolution for the Earth Land Surface Areas (CHELSA); topographic data (altitude) from USGS-NASA; habitat-type data from the Ecosystem Types of Europe based on the EUNIS (European Nature Information System) habitat classification. The habitat data was grouped into 11 level-1 habitat-type classes. Data on soil parameters was obtained from the German Federal Institute of Geoscience and Natural Resources (BGR), e. g. from the nationwide Soil Map 1 : 200,000 of Germany (BGR 1997–2018), except for soil moisture which was downloaded from the European Space Agency's (ESA) Climate Change Initiative (CCI) Soil Moisture product. To avoid autocorrelation and multicollinearity within these predictor variables, significant and independent predictor variables were selected during model calibration using principal component analysis (PCA), Pearson correlation and variance inflation factors (VIF) as well as AIC criteria during model building. This resulted in a final set of 11 environmental predictor variables on climate (average annual temperature and precipitation), habitat type and soil

(soil depth, soil pH, soil moisture, soil organic-matter content, texture [clay & silt content] and structure [bulk density, porosity]).

Four different model algorithms were tested for their abilities to predict earthworm distribution: two traditional regression models (Generalized Linear regression Models [GLMs], Generalized Additive Models [GAMs]) and two machine-learning algorithms (Generalized Boosting Models [GBM] and Random Forest models [RF]. The models were calibrated using the 11 predictor variables and – as response variables total density and species richness at the community level as well as occurrence probabilities at the species level. We ultimately used macro-ecological methods (MEM) to predict and map earthworm species richness and total community due to its advantages over stacked species distribution models (SSDM), in particular because data for rare species were very patchy resulting in underprediction of these metrics with SSDMs.

Model performance was assessed using a split-sample (cross validation) approach, whereby the data was split into training and test datasets at a 70:30 ratio. The training datasets were used to calibrate the models, while the test datasets were used to evaluate model predictive performance. Model goodness-of-fit was evaluated using the coefficient of regression determination (R^2) as well as the concordance index (C-index) for the quantitative response variables (community total density and species richness) and the threshold independent statistics “Area Under the receiver operating characteristic Curve” (AUCROC) and Kappa scores for the species-specific distribution-probabilities. Furthermore, for the quantitative response variables, plots and regressions of the observed-to-predicted values were calculated.

All model predictions (as spatial raster files) were imported to a GIS environment for visualisation (mapping) and further assessment. As a first approximation of earthworm potential diversity, a GIS “overlay” of the earthworm community total-density and species-richness modelling results was generated, producing a five-level scale of potential habitat suitability for earthworm diversity. Species’ range size were estimated from the model prediction results, quantifying for each evaluated species the extent of occurrence (EOO) and area of occupancy (AOO), which were used to categorize species into IUCN threat classes.

Finally, each predictor variable’s quantitative contribution to the model results was determined as well as the response variables’ (earthworm total density or species richness, as well species’ occurrence probability) response curves to each numerical environmental variable.

While calibrating the GLM models, a polynomial effect was very marginal when assessed by AIC. Therefore, for comparability, only the linear GLM model was retained. The GAMs were also subsequently dropped, as there was no significant difference in its performance compared to GLMs. Overall, the tested models’ prediction performance showed that GLM performance was very low. RF had the highest predictive performance for both community- and species-level models, while GBM goodness-of-fits were intermediary between GLM and RF. RF was therefore selected for all final predictions and mapping.

The models predicted earthworm community total density to range from 10 to maximally 800 ind. m^{-2} , with an average of 350 ind. m^{-2} per site. Species richness predictions ranged from 1 to 12 species, with an average of 3 species per site throughout Germany. Higher total community densities (> 400 ind. m^{-2}) were predicted especially in forests and grasslands in north-eastern Germany. However, in these regions, species richness was predicted to be relatively poor with an average of 2 species per site. An inverse relationship between species richness and total density is a common occurrence in ecology. The density/species-richness relationship often follows a negative slope, where areas may exhibit high individual densities, but low species richness. A region with less favourable environmental conditions may only allow occurrence of few species, but these in large populations (individual densities), perhaps due to reduced competition from

other species. A comparison with predicted species distributions revealed that in these areas primarily epigeic species such as *D. octaedra* (but also *L. rubellus*) were primarily responsible for the predicted high total densities.

Higher earthworm density and species richness were predicted for regions along the Rhine Valley, in Thuringia, and in Bavarian grasslands, with the average number of occurring species per site predicted to reach 10 and total density 300 ind. m⁻². The GIS overlay of earthworm diversity suitability showed especially the Rhine River valley, large areas in Thuringia, and the Bavarian Alps to be regions of potentially high earthworm diversity.

Three general patterns of spatial distribution ranges were identified in the model predictions of species occurrence probabilities: (1) species with large distribution ranges, (2) species with mid-range distributions, and (3) species with restricted or small distributional ranges, including endemic species or those limited to unique habitats. “Large-range” species in Germany consisted of two endogeic species (*A. caliginosa*, *A. rosea*) and one epigeic and anecic species each (*L. rubellus* and *L. terrestris*, respectively). The Red List of Germany lists all these species as being “very common”. These species were predicted to occur equally in many diverse habitat types, confirming reports documenting them as being ubiquitous or euryoecious, occurring in various habitat and land-use types.

The mid-range species, consisting of one endogeic (*A. chlorotica*), two epigeic species (*D. octaedra*, *L. castaneus*) and one anecic species (*A. longa*) were predicted to be mostly distributed in regional clusters, but not as broadly as the “large-range species”. The German Red List lists these all as being “common”. Although predicted to occur in many different habitat types (euryoecious), they appear to be more habitat discriminant, with preference optima in specific habitat types, which differ from species to species. Other observational reports of the habitat preferences of these species are generally in accord with the model predictions.

The species identified as “restricted-range” all showed higher occurrence probabilities limited to specific regions and habitat types. The Red List of Germany lists them all as being “rare” or “very rare”. Among the species predicted to have restricted distribution ranges were two epigeic species (*D. attemsi* and *A. eiseni*). *A. limicola*, as a habitat-specific species and known to occur in moist to wet sites, was predicted to show its strongest populations along the Rhine river, while *L. badensis*, a species endemic to south-western Germany, was predicted to have a very restricted and narrow distributional range in this region but also predicted, i.e., for the Bavarian Alps, where it has to date never been found. Nonetheless, the models predict a high potential habitat suitability for *L. badensis* in the Alps, despite its observed absence, underscoring its status as endemic to southwest Germany.

According to the IUCN criteria for classifying species’ threat status (E00 and AOO), none of the evaluated species were considered to be critically endangered. Four species with extensive E00 and AOO sizes (*A. caliginosa*, *A. rosea*, *L. terrestris*, and *L. rubellus*) cannot be considered to be threatened. Based on AOO assessment, three species can be classified as being vulnerable (*A. longa*, *D. octaedra*, *L. castaneus*) and three species (*A. eiseni*, *A. limicola*, *D. attemsi*) as endangered. The occurrence data for *L. badensis* was not sufficient to calculate geographic range sizes.

The environmental variables contributing most to the predictions of earthworm total density were soil moisture, habitat type, and average annual precipitation. The remaining predictors for total density contributed less, but fairly equally to the model results, with the exceptions of average annual temperature and soil pH, which accounted for less than 5%. The environmental contributions to predicted species richness were dominated by clay content and habitat type. Climate variables (average annual precipitation and temperature and pH contributed moderately

to the species-richness predictions, while the remaining variables equally contributed slightly. Soil depth played the most minor role in the species-richness predictions.

Density predictions increased below 500 mm a⁻¹ total annual precipitation, but reduced above this threshold. Average annual temperature did not affect density predictions below 10 °C, but increases were predicted above 10 °C. These climate parameters had an inverse effect on species-richness predictions, which increased above 500 mm a⁻¹ precipitation and decreased between 6 °C and 10 °C average annual temperature. Increasing soil moisture led to lower total density predictions, but did not strongly influence the species-richness projections.

Density predictions dramatically reduced at about pH 4, only to increase again above pH 5. The predictions of species richness, on the other hand, remained very low below pH 4, and strongly increased above this value. The density predictions increased above SOM contents of 6 %, while showing little influence to species-richness predictions. An influence of soil texture was best represented by clay content, where predictions of both total density and species richness increased at high clay contents. Soil structure also influenced the earthworm community predictions, whereby density projections strongly increased at a total porosity above 15 %, but species-richness projections decreased continuously with increasing porosity. Soil bulk density apparently negatively affects earthworm communities, as both total-density and species-richness predictions abruptly and dramatically decreased above bulk densities of 1.5 g cm⁻³.

Previous studies using simpler statistical methods also identified climate (average annual precipitation and temperature) as the almost exclusive driver of earthworm communities (total density, species richness) at global or continental scales. While our study confirmed the combined role of temperature, precipitation and soil moisture, it also identified, i.e., habitat type, soil pH and soil organic matter as important drivers. The relevant drivers of earthworm biodiversity are apparently scale dependant; climate parameters being important at global and (sub-) continental scales, while habitat and soil factors become more important at smaller spatial scales.

Regarding the species-specific models of occurrence probability, total annual precipitation and the related soil moisture accounted for up to 25 % or more of the predictions in many species and were found to be essential predictor variables for ¾ of the tested species. The occurrences of *D. attemsi*, *A. eiseni*, *L. rubellus* and *A. chlorotica* were predicted to also be generally highly dependent on climate (average annual precipitation and temperature), while the occurrence probability of *D. octaedra* was predicted to be highly related to annual average temperature. Habitat type was the next environmental predictor contributing importantly to species' occurrence probabilities. Only for *A. eiseni* was habitat type less important (7 %). The contribution of soil pH to the occurrence predictions was moderate across all species. *A. longa* and *A. caliginosa* were predicted to be highly dependent on soil pH. Soil organic matter contributed more strongly to the predictions of epigeic species and endogeic species such as *A. limicola* and *A. caliginosa*, compared to much lower prediction contributions for the majority of endogeic and anecic species. Soil texture (clay and silt content) had comparatively less influence on most species' occurrences, except for *A. longa*, *D. octaedra* and *L. terrestris*, which were moderately influenced by soil texture. Soil structure (porosity and bulk density) in general only moderately influenced model results, although occurrence probabilities of species such as *L. rubellus* and *L. badensis* were highly influenced by bulk density and *L. castaneus* and *A. rosea* by porosity. Soil depth contributed least to the model results.

Climate parameters (average annual temperature and precipitation) and the related soil moisture presented different patterns of influence on the occurrence predictions of epigeic and endogeic species. Whereby the occurrence probabilities of epigeic species generally decreased with increasing annual mean temperature and decreasing mean annual precipitation and soil

moistures, the occurrence predictions of endogeic species often increased with higher annual mean temperature and precipitation as well as soil moisture. The precipitation thresholds at which occurrence probabilities increased were species specific.

Species' prediction responses to soil pH were also different between epigeic and endogeic/anecic species. The occurrence probabilities of epigeic species were generally larger at lower pH values, while that of endogeic and anecic species were larger at higher pH values. Interestingly, the threshold at which occurrence probabilities either increased or decreased were generally around pH 4.0. The predictions thus identified acidophobus and acidophilus (or -tolerant) species, which confirm species classifications given by previous authors, who also described a threshold of pH 4 - 5. No consistent response to soil organic matter was observed among life-form groups, and the prediction responses were instead species specific. While some species showed reduced occurrence probabilities with increasing SOM content (with optima < 6 %), other species exhibited positive responses to increasing SOM (with maximum probabilities at 8 – 12 % SOM). These results corroborate earlier descriptions of species-specific preferences of organic matter contents.

Soil texture (included into the models as clay and silt content) only contributed substantially to the occurrence predictions of a few species, and were not related to earthworm life-form. For instance, the predicted occurrence probabilities were larger at both higher clay and silt contents for *A. eiseni*, *A. rosea*, *A. caliginosa* and *L. terrestris* as did those of *A. limicola* at higher clay content suggesting a preference for finer textured soils. Contrarily, the responses of *L. rubellus*, *D. attemsi*, *A. chlorotica* and *A. longa* generally decreased with higher clay and silt content – suggesting preferences for coarser soils. Differences between life-form types were observed concerning responses to soil structure. Most epigeic species exhibited higher occurrence probabilities at higher levels of porosity and lower bulk densities, suggesting a preference for looser soil. Contrarily, the predicted responses of some endogeic species and the anecic *L. terrestris* were larger at lower levels of porosity and higher bulk densities, suggesting a preference for denser soils.

In conclusion, this study is to the best of our knowledge the first comprehensive analysis modelling earthworm distribution at a national scale, including the most important species and differentiating among multiple environmental drivers and the predicted responses to each. The earthworm distribution models provided an effective approximation of earthworm distribution and its drivers in Germany. Such information is invaluable for future scientific field studies and a prerequisite for soil-biodiversity monitoring programs, which require standardized baseline and threshold values for result assessment. A tool is currently being developed to extract reference values from the model results based on specific site conditions, explicitly for use in soil-biodiversity monitoring programmes. While such programmes will help validate the model results, we call for wide-spread recording of environmental (especially soil) parameters concomitantly with biodiversity surveys, in order to improve the thematic association between species and environmental drivers and, thereby, model precision. Attention should be given particularly to species with restricted ranges, such as *D. attemsi*, *A. eiseni*, and *L. badensis*. For species with clearly defined habitats, such as *A. limicola* in wetlands and *D. octaedra* in forests, the habitats in which they can be found should be monitored for possible habitat degradation. We further suggest detailed studies on the endemic *L. badensis*, which would allow more precise SDMs and calculation of geographic range sizes, providing a better assessment of its realised distribution and assessment needs.

1 Einführung

1.1 Hintergrund

Einen großen Teil der Biodiversität terrestrischer Ökosysteme macht die bodenbewohnende Fauna aus (FAO 2020). Zum Beispiel kann ein Quadratmeter Waldboden bis zu 1000 Arten mit mehreren 100 000 Lebewesen Raum geben (Anderson, 1975). Laut Behan-Pelletier und Bisset (1992) kann ein Waldboden – nach den Korallenriffen – die größte phylogenetische Artendiversität aller Habitate der Erde aufweisen. André *et al.* (1994) bezeichnete den Boden – zusammen mit den Korallenriffen und dem tropischen Regenwald – als das letzte unentdeckte Gebiet in der Biodiversitätsforschung. Die Bodenbiodiversität stellt folglich ein wichtiges kulturelles Gut dar, dessen Schutz und Erhalt nicht allein aus seinem Wert an sich oder allgemeinen ethischen Überlegungen erforderlich ist. Die Bodenbiodiversität ist auch unverzichtbar für die Erfüllung der Bodenfunktionen, die das Funktionieren der terrestrischen Ökosysteme sicherstellen. Der Großteil der Ökosystemdienstleistungen (*ecosystem services*) terrestrischer Habitate beruht auf Bodenfunktionen (Adhikari & Hartemink, 2016), wobei praktisch wiederum alle Bodenfunktionen sich aus Prozessen ableiten, an denen Bodenorganismen maßgeblich beteiligt sind (Gardi & Jeffery, 2009; Turbé *et al.*, 2010). Zum Beispiel spielt Bodenbiodiversität eine wichtige Rolle bei der Zersetzung von organischem Material. 90 % der gesamten natürlichen terrestrischen Nettoprimärproduktion (NPP) werden im Boden recycelt und dort abgebaut (Bardgett, 2005), mit entsprechenden Wirkungen auf den Nährstoffkreislauf, den Erhalt der Bodenstruktur, die Wasserinfilt ration und -reinigung, die Schädlingsbekämpfung, den Abbau von toxischen Substanzen, usw. (Brussaard, 1998; Lavelle *et al.* 2006; Briones, 2018; Delgado-Baquerizo *et al.*, 2020). Damit ist die Primärproduktion von Nahrung und Fasern eng verbunden mit Prozessen, die von Bodenorganismen gesteuert werden (z. B. Wardle, 2002; Brussaard *et al.*, 2007; Bakker *et al.*, 2019).

Die ungeheure Größe der Bodenbiodiversität und ihre funktionale Wichtigkeit ist seit Jahrzehnten bekannt, und das Abkommen von Rio 1992 bestätigte zum ersten Mal die Notwendigkeit des Schutzes von Bodenlebewesen (André *et al.* 1994; Giller 1996; Hagvar, 1998). Jedoch ist diese Tatsache erst seit kurzem von Politik und Verantwortlichen entsprechend wahrgenommen worden (Turbé *et al.*, 2010; FAO *et al.* 2020), wie die jetzt erfolgte Aufnahme der Biodiversität in die EU-Bodenschutzstrategie (Eur. Comm. 2021) zeigt. Dies gilt auch für die überarbeitete Gemeinsame Agrarpolitik (EU(2001)2115, EU(2001)2117) sowie die Aufnahme eines Bodenmonitoring-Zentrums in die Koalitionsvereinbarung¹ der deutschen Bundesregierung.

Bodenschutz erfordert koordinierte Schritte zur Bewertung und zum Erhalt der Bodenbiota (Dunbar *et al.*, 2013). Diese Schritte beinhalten ein immer wichtiger werdendes Verständnis der spezifischen Verbreitung von Bodenbiodiversität für ein effektives Bodenmanagement und dessen Bewertung (Boumal & Montanarella, 2016; EASAC 2018).

Eine wichtige Grundlage für den Bodenschutz stellen Monitoringprogramme dar, bei denen Bodenparameter auf Bodendauerbeobachtungsflächen regelmäßig erfasst werden. Eine Vielzahl von europäischen Richtlinien fordern ausdrücklich ein Bodenmonitoring, so z. B. die Gemeinsame Agrarpolitik (CAP, z. B. EU/2021/2116, Art. 25, 66, 67, 70 usw.); die Verordnung über Industrieemissionen (EU Com 2010/75/EU; Einleitung § 23, Art. 14, 16), die Klärschlammverordnung (EU Com 86/278/EC; Einleitung, Anhang II), die Biodiversitätsstrategie (COM(2011)244; Art. 2.3), der Forstaktionsplan (COM(2006)302; Art. 3.2) usw. Die EU-Bodenstrategie für 2030 widmet dem Bodenmonitoring ein gesamtes Kapitel und hebt darin besonders die Bodenbio-

¹ <https://www.bundesregierung.de/resource/blob/974430/1990812/04221173eef9a6720059cc353d759a2b/2021-12-10-koav2021-data.pdf?download=1>

diversität hervor (COM(2021)699, Kap. 5). Innerhalb Deutschlands fordert das Bundesbodenschutzgesetz (BBodSchG 1998) ein Bodenmonitoring, besonders hinsichtlich Bodenschutzmaßnahmen und Bodensanierung (z. B. §10(1)). Die Einrichtung von Bodendauerbeobachtungsflächen ist Aufgabe der Bundesländer (§21 Abs. 4). Das BBodSchG begründet nur die Notwendigkeit, „biologische Charakteristika“ von Böden zu schützen und zu bewahren (und spezifizierter: „das Habitat für Bodenorganismen“, §2(2)). Ein spezifischer Schutz der Bodenbiodiversität wird im Hinblick auf eine Revision dieses Gesetzes erwartet, wie in den Koalitionsvereinbarungen zur Bildung der deutschen Bundesregierung festgelegt. Ebenfalls wird ein permanentes Bodenmonitoring in der das BBodSchG begleitenden Bundes-Bodenschutz- und Altlastenverordnung (BBodSchV, §5(3), §10(1), Anhang 3) gefordert, wobei hier hauptsächlich die Wiederaufbereitung sowie Vorsorge- und Schutzmaßnahmen gegen Verschmutzungen und deren Folgen im Vordergrund stehen. Um dieses Ziel zu erreichen, sind in Deutschland zahlreiche Bodenmonitoringprogramme etabliert worden. Das größte umfasst etwa 800 Bodendauerbeobachtungsflächen, die über ganz Deutschland verteilt sind (UBA 2002). Obwohl vorrangig physikalisch-chemische Bodenparameter (und meist mikrobielle Charakteristika) auf diesen Flächen überwacht werden (vgl. Kaufmann-Boll *et al.*, 2011), beobachten zahlreiche Bundesländer auch Bodentiere, insbesondere Regenwürmer (z. B. Sachsen-Anhalt, Schleswig-Holstein, Brandenburg). Die deutsche Bodenzustandserhebung überwacht die aktuelle Bodensituation von landwirtschaftlichen² und forstwirtschaftlichen Böden³. Obwohl hier im Allgemeinen die physikalisch-chemischen Eigenschaften des Bodens untersucht werden, wird zurzeit die Aufnahme der Bodenbiodiversität diskutiert. Das Bundesministerium für Landwirtschaft und Ernährung (BMEL) hat vor kurzem ein nationales Monitoringprogramm zur Biodiversität von landwirtschaftlichen Flächen (MonViA⁴) beschlossen, in dem auch die Aufnahme der Bodenfauna (Schwerpunkt Regenwürmer, aber auch andere Bodentiere) ausdrücklich geplant ist. U. a. haben diese Programme dazu beigebracht, dass das kürzlich eröffnete deutsche Monitoringzentrum für Biodiversität⁵ seine erste Expertenkommission beauftragt hat, Empfehlungen für ein deutschlandweites Bodenbiodiversitätsmonitoring zu koordinieren.

In den letzten Jahren wurden auch in der Wissenschaft die Notwendigkeit eines Bodenbiodiversitätsmonitorings beschrieben und Ansätze dazu geliefert (z. B. Gardi *et al.*, 2009, Cluzeau *et al.*, 2012, Pulleman *et al.*, 2012, Griffiths *et al.*, 2016, Orgiazzi *et al.*, 2016, van Leeuwen *et al.*, 2017, Guerra *et al.*, 2021). Während diese Arbeiten die generellen Anforderungen, Ansätze und allgemeinen Konzepte zum Bodenbiodiversitätsmonitoring aufzeigen, sind in den letzten zwei Jahrzehnten konkrete nationale Programme entwickelt worden, in denen detailliert Ansätze zur Bewertung gemacht werden; so z. B. in England die „SoilPacs“ (Weeks, 1998), die deutsche Bodenbiologische Standortklassifizierung (BBSK, Römbke *et al.*, 2000, 2002) oder der holländische Indikator zur Bodenqualität *Biological Indicator for Soil Quality* (BISQ, Rutgers *et al.*, 2008, 2009), um nur einige zu nennen. Alle Programme basieren auf Auswertungen der Monitoringergebnisse und einem Vergleich mit Referenzwerten, wobei die Referenzwerte selbst aus Geländeuntersuchungen ausgewählter Referenzstandorte stammen. Jedoch ist es ohne grundlegende verfügbare Daten zur Bestimmung von Referenzwerten für die Bodenbiodiversitätsbewertung sehr schwierig, wenn nicht gar unmöglich, wirksam solche Ziele auf nationaler Ebene zu erzielen (Glaesner *et al.*, 2014). Die Europäische Behörde für Lebensmittelsicherheit (EFSA) tritt für die getrennte Ermittlung von Hintergrundwerten für landwirtschaftlich genutzte und für natürliche Flächen

² <https://www.thuenen.de/en/ak/projects/agricultural-soil-inventory-bze-lw/>

³ <https://www.thuenen.de/en/wo/projects/soil-protection-and-forest-health/projekte-bodenzustandserhebung/national-forest-soil-inventory/>

⁴ <https://www.agrarmonitoring-monvia.de/monvia/>

⁵ <https://www.monitoringzentrum.de/>

ein (Ockleford *et al.*, 2017). Sachgemäßes und verfügbares Wissen über Bodenbiodiversität, mit dem die Folgen von Landnutzung, Pflanzenschutz und Bodenbewirtschaftung beurteilt werden können, ist entscheidend für die Bewertung und zum Erhalt der Bodenbiodiversität sowie einer angemessenen nationalen Bodenschutzpolitik. Jedoch gibt es weder EU- noch bundesweit adäquate Richtwerte zur Bewertung der vorliegenden Biodiversitätsdaten. Während die Bodenbiodiversitätsbeobachtungen innerhalb des deutschen o. g. Monitoringprogramms natürlich diese Daten generieren könnten, würde es Jahre erfordern, Referenzwerte für alle Regionen, Habitate, die verschiedenen Landnutzungstypen, Klimazonen, Bodentypen usw. Deutschlands zu bekommen. In der Tat ist der meist genannte Hinderungsgrund für das Arbeiten mit den Werten aus dem Bodenbiodiversitätsmonitoring in den aktuellen Programmen das Fehlen von Referenzwerten für den Vergleich der Ergebnisse (pers. comm. R. Baritz, EEA).

Ein weiterer international gebräuchlicher Ansatz in den Biodiversitäts- und Artenschutzprogrammen besteht darin, Zusammenhänge zwischen Arten bzw. Lebensgemeinschaften und Umweltparametern statistisch zu modellieren, um diese wiederum dafür zu nutzen, potentielle Vorkommen hochzurechnen, und zwar basierend auf der räumlichen Ausprägung der Umweltparame-*ter (species distribution models – Artenverbreitungsmodelle, Guisan *et al.*, 2017)*. So ist es z. B. möglich, Vorkommen von Arten bzw. Lebensgemeinschaften auf regionaler bzw. landesweiter Ebene vorherzusagen. Derartige Kartiermethoden bieten einen grundsätzlichen Ansatz zur Entscheidungsunterstützung für politische Bemühungen zum Erhalt der Biodiversität und des Ökosystemdienstleistungen in der EU (Maes *et al.*, 2012). Die Notwendigkeit von verlässlichen Informationen zur Verbreitung der Lebewesen, die Ökosystemdienstleistungen erbringen, ließ die EU-Arbeitsgruppe *Mapping and Assessment of Ecosystems and their Services* (Kartierung und Bewertung von Ökosystemen sowie ihrer Dienstleistungen, MAES⁶) entstehen. Zwar werden im Rahmen dieser Initiativen weder die tatsächliche Bodenbiodiversität untersucht, noch stellen sie Daten zur Bodenbiodiversität zusammen. Solche kleinmaßstäbigen Karten zur voraussichtlichen Verbreitung der Bodenbiodiversität stellen hingegen einen großen Schritt in der Formulierung von landesweiten bzw. standortspezifischen Referenzwerten für die Bewertung der Monitoringergebnisse dar.

Bodenbiodiversität wird nicht nur, wie oben erwähnt, durch tausende von Arten und Millionen von Individuen charakterisiert, sondern diese gehören darüber hinaus zu dutzenden von taxonomischen Hauptgruppen. Ein Monitoring bzw. eine Modellierung von allen vorkommenden taxonomischen Gruppen durchzuführen ist weder vom Ressourcenaufwand zu rechtfertigen noch effizient. Daher gilt es, eine Auswahl zu treffen. Das ENVASSO-Projekt⁷ des 6. europäischen Forschungsrahmenprogramms hat Indikatoren für die Bodenbewertung, einschließlich der Bodenbiodiversität, innerhalb der Monitoringprogramme ausgearbeitet. Dieses Konsortium hat insbesondere die Regenwürmer (*Lumbricidae*) und Springschwänze (*Collembola*) identifiziert, die den größten indikativen Wert für das Level-1-Monitoring hinsichtlich einer großen Anzahl von Habitaten und Landnutzungstypen aufweisen (Huber *et al.*, 2008). In dem darauffolgenden 7. Forschungsrahmenprogramm hat das EU-Projekt EcoFinders⁸ die Biodiversität dieser Gruppen innerhalb einer weiten Skala von Bodentypen, Klimazonen und Landnutzungen europaweit charakterisiert. Obwohl eine große Anzahl von Gruppen an Bodenorganismen bewertet wurden, stellten die Regenwürmer eine Schlüsselgruppe dar (z. B. Faber *et al.*, 2013). Innerhalb Deutschlands hat die landesweite Edaphobase-Bodendiversitätsstudie (sog. Länderstudien, Toschki *et al.*, 2021) die Artengemeinschaften von ausgewählten Bodenfaunagruppen, hier auch die Regenwürmer, in zahlreichen Regionen Mittel- und Süddeutschlands untersucht. Alle Studien haben

⁶ <https://biodiversity.europa.eu/maes>

⁷ <https://esdac.jrc.ec.europa.eu/projects/envasso>

⁸ <https://projects.au.dk/ecofinders>

eine große Anzahl von Daten zur Verbreitung von Regenwürmern generiert. Die o. g. deutschen Monitoringprogramme haben ebenfalls Daten über Regenwurmgemeinschaften in vielen Boden-dauerbeobachtungsflächen in Deutschland geliefert. In mehreren Studien ist der indikative Wert von Regenwürmern hinsichtlich der Bodenbewertung untersucht worden (z. B. Krück *et al.*, 2006, Fründ *et al.*, 2011, Spurgeon *et al.*, 2013, Crittenden *et al.*, 2014, Kanianska, 2016). Darüber hinaus ist im Gegensatz zu anderen Gruppen von Bodenlebewesen der Regenwurm einfach zu untersuchen. Das Vorhandensein von anwenderfreundlichen Artenbestimmungsbüchern (z. B. Krück, 2018) sowie von taxonomischen Experten in ganz Deutschlands machen zusammen mit den Konsortien und den o. g. Studien den Regenwurm einzigartig für die Formulierung von deutschlandweiten Referenzwerten für das Bodenbiodiversitätsmonitoring. Das Erstellen von Karten zur potenziellen Verbreitung ist ein erster Schritt in Richtung der benötigten Referenzwerte, der auf Grundlage der Regenwurmbeobachtungen in den laufenden Monitoringprogrammen ermöglicht werden kann. Die tragende Rolle des Regenwurms, z. B. für die Verbesserung der Bodenstruktur und der Bodenqualität, ist schon länger bekannt (z. B. Pérès *et al.*, 1998; Suthar, 2009; Hedde *et al.*, 2103; Amossé *et al.*, 2015; Blouin *et al.*, 2015). Jedoch gibt es noch keine vollständigen quantitativen und räumlichen Aussagen, wie sich diese Bodentiergruppe bei umweltbedingten Schwankungen bzw. im größeren Raum verhält und wodurch Änderungen ausgelöst werden.

In den letzten Jahren ist die Entwicklung von Artenverbreitungsmodellen (SDMs – *species distribution models*) als mächtiges Werkzeug zum Verständnis der Biodiversität, ihrer Verbreitung und den möglichen Triebkräften, die zu dieser Verbreitung führen, immer wichtiger geworden. Drei Phasen in der Entwicklung der *species distribution* (SD) sind hier zu nennen: (i) nicht-räumliche statistische Quantifizierung von Arten-Umwelt-Beziehungen basierend auf (normalerweise studienspezifischen) empirischen Daten; (ii) Experten-basierte (nicht statistische, nicht empirische) räumliche Beschreibungen von Artenverbreitungen und (iii) räumlich explizite statistische und empirische Modellierungen von Artenverbreitungen (SDM) (Guisan and Thuiller 2005; Ferrier *et al.*, 2007). Der letzte Punkt beinhaltet die Kombination von statistischen Modellierungen mit raumbezogenen Daten mit dem Ziel, aufzudecken, wie sich Umweltvariablen auf die Artenverbreitung über Raum und Zeit auswirken (Elith *et al.*, 2006). Zu diesem Zeitpunkt bezog sich die Biodiversitätsmodellierung vorrangig auf größere Pflanzen und größere (Wirbel-)Tiere; es gab nur wenige Studien für wirbellose Tiere und nahezu keine für wirbellose Bodentiere.

SDM ist auch bekannt als Habitateignungsmodellierung und wird in der Ökologie, den Umweltwissenschaften und der Biogeographie dafür genutzt, potentielle und tatsächliche Artenverbreitungen zu bewerten, sei es zum Zweck des Erhalts von bedrohten und gefährdeten Arten oder zur Kontrolle von invasiven Arten (Kumar und Stohlgren 2009; Urbani *et al.*, 2017; Wang *et al.*, 2018). SDM ist ein wirksames Werkzeug zur Identifizierung von aussagekräftigen und wissenschaftlich belastbaren Beziehungen zwischen dem Vorkommen von Arten und Umweltvariablen (meist bioklimatische Variablen), die potentielle Habitate einer Zielart zu identifizieren helfen. SDM beschränkt sich nicht auf die Vorhersage von momentanen Artenverbreitungen, sondern spiegelt auch die Muster der Artendiversität in Zeit und Raum wider. Die Sorge der Wissenschaft hinsichtlich der Auswirkungen des Klimawandels auf Umwelt und Natur auf lokaler, regionaler und globaler Ebene hat zur Anwendung von SDM als wirksames Werkzeug für das Verständnis der zukünftigen Entwicklung der Umweltvariablen und in ihrer Folge der Biodiversität auf diesen Ebenen (Brambilla *et al.*, 2014; Hu *et al.*, 2015).

Artendiversität ist eine Funktion der Zusammensetzung von Organismengemeinschaften (*welche* Arten vorkommen), des Artenreichtums (*wie viele* Arten sympatrisch auftreten), der Gesamtabundanz und der Abundanz der einzelnen Arten sowie der räumlichen Muster ihrer Vor-

kommen. Alle diese Aspekte müssen in die Abschätzung der Biodiversität einfließen. Nur einzelne dieser Parameter zu betrachten, reicht für eine Biodiversitätsbewertung innerhalb einer Gemeinschaft nicht aus, weil sich z. B. zwei Gemeinschaften in ihrem Artenreichtum gleich sein können, aber unterscheiden in Dichte und Artabundanzen (Groves, 2022). Die valide Prognose der potentiellen Verbreitung der Bodenbiodiversität verbunden mit der Zusammensetzung, der Abundanz und dem Artenreichtum der Organismengemeinschaft ist eine elementare Voraussetzung für die Bewertung, den Erhalt und den Schutz der Biodiversität sowie der daraus resultierenden Bodenfunktionen.

Es existieren mehrere SDM-Modellierungsverfahren bzw. -Algorithmen, wie z. B. Generalised Linear Models (GLM), Classification Tree Analysis (CTA), Multivariate Adaptive Regression Spline (MARS), Maximum Entropy (MAXENT), Random Forest (RF) und Generalised Boosted Regression Model (GBM), die alle Stärken und Schwächen haben (Li and Wang, 2013; Valavi *et al.*, 2021). Einige davon sind sehr empfindlich gegenüber der Probengröße, was sich sehr auf ihre Fähigkeit zur Vorhersage der räumlichen Verbreitung der Arten und Gemeinschaften auswirkt (Kumar and Stohlgren, 2009). Modellvergleich ist eine von zwei Methoden, modellabhängige Unsicherheiten in SDM zu überwinden. Die andere besteht darin, Ensemblemethoden zu verwenden, bei denen mehrere Modellierungsansätze kombiniert werden (Marmion *et al.*, 2008). Guisan *et al.* (2017) haben verschiedene Bewertungsparameter identifiziert, die eine Entscheidung für die beste Modellierung hinsichtlich der Voraussage der räumlichen Verbreitung ermöglichen. Diese gehen u. a. von konventionellen Anpassungsgüte-Statistiken wie R^2 (Root-Mean-Square-Abweichung, RMSE) über die schwellenwertunabhängige Fläche unter der "Receiver Operating Characteristic"-Kurve (AUC_{ROC}) bis hin zu Kappa-Statistiken (beobachtete vs. erwartete Prädiktorübereinstimmung, korrigiert um Stochastizität) und anderen. Die Wahl der für die Bewertung der Modellleistung zu verwendenden Metriken sollte jedoch von der Art und den Zielen der Studie sowie der Art der verwendeten Daten bestimmt werden (Guisan *et al.*, 2017, Zurell *et al.*, 2020).

In der Diskussion über Artendiversität in Ökologie und Biogeographie beziehen sich die englischsprachigen Ausdrücke *species range size* (Gebietsgröße einer Art) und *species distribution range* (Verbreitungsgebiet einer Art) beide auf das Areal einer Art und werden oft synonym verwendet, obwohl sie sich in Methodik und Zweck unterscheiden. Die *species distributional range* beschreibt das Vorkommen einer Art oder ihre Anordnung in einem geographischen Raum, ohne notwendigerweise seine Größe oder seine Ausdehnung zu bestimmen. Die *species range size* dagegen beschreibt die Fläche, auf der eine Art angetroffen wird, normalerweise angegeben in km^2 , und wird als Indikator für den Bedrohungsstatus einer Art verwendet. Im Gegensatz zu Arten mit einem größeren Verbreitungsgebiet sind Arten mit einem eingeschränkteren Verbreitungsgebiet vulnerabler und es besteht die hohe Wahrscheinlichkeit, dass die Art (örtlich) ausstirbt (Gaston & Fuller, 2009). Karten zur räumlichen Artenverbreitung können meist für die Bestimmung der Größe des Artenverbreitungsgebiets genutzt werden (Gaston & Fuller, 2009), und das eine kann zur Begründung des anderen herangezogen werden (Pulliam, 2000). Die Klassifizierung von Arten in verschiedene Bedrohungskategorien geschieht meist unter Berücksichtigung von geographischen Gebietsgrößen (Sheth *et al.*, 2020) und ist anerkannter und gelebter Standard bei der Erstellung der Roten Listen bedrohter Arten der IUCN, Punkt B1 and B2 (IUCN, 2001, 2012 b und 2022).

Zwei Parameter haben sich für die Bestimmung des geographischen Gebiets einer Art etabliert: die Ausdehnung des Vorkommens (EOO, *extent of occurrence*) und der Lebensraum (AOO, *area of occupancy*). Die von der IUCN definierte EOO ist „das Gebiet innerhalb der kleinsten durchgehenden vorstellbaren Grenzlinie, die um tatsächliche, angenommene oder vorhergesagte aktuelle Vorkommen eines Taxons gezogen werden kann, ohne vorübergehende Zufallsansiedlungen“

(IUCN 2001; 2012b). Die *area of occupancy* ist ein skaliertes Parameter, der das Areal des *geeigneten* Habitats angibt, das ein Taxon aktuell besetzt. Die AOO wird aus zwei Gründen in die IUCN-Kriterien aufgenommen: Zum einen ist AOO ein Maß für den "Absicherungseffekt" (Keith *et al.* 2018), der sich so erklärt, dass Taxa, die an vielen Stellen in oder über große Flächen in einer Landschaft vorkommen, gegen Risiken aus konkreter räumlicher Bedrohung "abgesichert" sind. Zum anderen korreliert AOO auch mit der Populationsgröße und kann so die Widerstandsfähigkeit einer Art hinsichtlich stochastischer (Aussterbens-)Ereignisse annähernd beschreiben (IUCN Standards and Petitions Committee, 2022).

Auch wenn beide Maße (AOO und EOO) verwendet werden können, um den Bedrohungsstatus von Arten einzuschätzen, sind Ansatz und Prinzip nicht der- bzw. dasselbe (Gaston & Fuller, 2009). Während die EOO grundsätzlich die gesamte geografische Ausdehnung des Vorkommens einer Art misst, einschließlich der Gebiete, die sich als ungeeignet für die Art erwiesen haben, um die Faktoren zu verstehen, die die Verbreitungsgrenzen einer Art bestimmen, und um den Grad der räumlichen Streuung der Risiken über die geografische Verbreitung einer Art zu bewerten (IUCN 2012), misst die AOO das tatsächliche geografische Gebiet, in dem eine Art vorkommt, und ist in den meisten Fällen gleich groß oder kleiner als die EOO-Werte (Hurlbert & Jetz, 2007).

Minimumkonvexpolygone (*minimum convex polygons*, MCP) bzw. konvexe Hüllen sind in der Fachwelt weitreichend anerkannt und wurden von der IUCN als geeignete Methode zur Berechnung der EOO von Arten übernommen. Sie sind definiert als das kleinste Polygon, in dem keiner der inneren Winkel größer als 180 Grad ist und alle Vorkommen einer Art enthält. Die AOO kann dagegen entweder durch ein Artenverbreitungsmodell oder die Darstellung der Daten zum Vorkommen auf einem 2 x 2 km-Raster bestimmt werden (Gaston & Fuller 2009; IUCN 2022).

1.2 Stand der Forschung in Deutschland

In einer Reihe von Studien wurde versucht, die Verbreitung von Regenwurmarten in Deutschland zu beschreiben und Umweltfaktoren für ihre Verbreitung zu identifizieren. Jänsch *et al.* (2013) fassen den aktuellen Wissensstand über die wichtigsten in Deutschland vorkommenden Regenwurmarten, ihre bekannten Verbreitungsgebiete und die wichtigsten Lebensraum- und Bodenparameter, die ihr Vorkommen beeinflussen, zusammen. Diese Monographie basiert auf den in Edaphobase verfügbaren Daten und übertrug die Fundorte der Arten aus dieser Datenbank in eine Karte. Außerdem wurden mit den Fundorten in der Datenbank verknüpfte Umweltdata verwendet, um die Häufigkeit des Vorkommens mit z. B. Habitattypen und einigen Bodenparametern in Beziehung zu setzen, aber es wurden keine externen Umweltdaten oder fortgeschrittene statistische Modellierung verwendet. Lehmitz *et al.* (2016) führten eine umfangreiche deskriptive Studie über Regenwürmer durch und erstellten eine Artenliste und die geografische Verteilung der Regenwurmarten in Deutschland. Obwohl diese Arbeit sehr gute, aktuelle allgemeine und beschreibende Informationen darüber liefert, wo verschiedene Regenwurmarten bisher gefunden wurden, eignet sie sich nicht zur Abschätzung der potentiellen Verbreitung der Arten. Eine aktuelle Monographie über deutsche Regenwürmer (Krück, 2018) beschreibt zwar textlich die allgemeine Verbreitung und Habitatabhängigkeit einzelner Arten und ist hinsichtlich der vorgestellten Arten sehr ausführlich, beschränkt sich aber auf Nordostdeutschland und basiert primär auf Expertenwissen. Rutgers *et al.* (2016) haben die Verteilung von Regenwurm-Gemeinschaftsparametern auf kontinentaleuropäischer Ebene anhand ausgewählter nationaler Datensätze und eines verallgemeinerten linearen Modells (GLM) statistisch modelliert. Diese Studie war eher ein Proof-of-Concept für die Modellierung der Verteilung von Bodentieren und modellierte nicht die Verteilung einzelner Arten. Es gibt bisher noch keine Studie für Deutsch-

land, in der die Verbreitung bestimmter Regenwurmarten mit Hilfe mehrerer Modellalgorithmen modelliert und räumlich kartiert wurde. Die vorliegende Studie zielt daher auf die statistische Modellierung der potenziellen standortbezogenen Regenwurmverbreitung in Deutschland ab, und zwar sowohl für allgemeine Gemeinschaftsparameter (Artenreichtum und Gesamtbundanz) als auch für das Vorkommen einzelner Arten. Da die Datenerhebung für die Rote Liste der Regenwürmer in Deutschland (Lehmitz *et al.*, 2016) umfassend war und alle Daten in Edaphobase aufgenommen wurden, basiert die aktuelle Studie auf allen Regenwurmfunddaten, die im Edaphobase-Portal über einen langen Zeitraum und an vielen Standorten gesammelt wurden, sowie auf Umweltdaten auf Standortebene, die gleichzeitig mit den Daten zum Regenwurmvorkommen erhoben wurden. Fehlende und zusätzliche Umweltdaten wurden aus externen Quellen ergänzt. Die verwendeten Modellierungstechniken beruhen in erster Linie auf den (korrelierten) Anpassungen der Arten an ihre lokale Umgebung (korrelative Artenverteilungsmodelle, SDMs).

1.3 Ziel

Das allgemeine Ziel dieser Studie ist, eine Karte zur Verbreitung von Regenwurmgemeinschaften und -arten räumlich zu modellieren zum Zweck der visuellen Interpretation und schnellen Bewertung des Verbreitungsstatus. Hierzu sollten:

- ▶ Die Anwendung von korrelativen Modellierungstechniken zur Kartierung der räumlichen Verbreitung und der geographischen Ausdehnung von Regenwurmgemeinschaften sowie sorgfältig ausgewählten Regenwurmarten in Deutschland geprüft und
- ▶ die Gewichtung von Umweltprädiktoren auf der Grundlage ihrer Signifikanz in den korrelativen Modellierungen zu identifizieren und bestimmen.

Primär basieren die Modellierungstechniken auf den entsprechenden Artenanpassungen an die lokalen Umweltbedingungen (korrelative Artenverbreitungsmodelle, *correlative species distribution models*, SDM).

Die Ergebnisse sollen Hintergrundwerte für einen Vergleich mit zukünftigen Monitoringergebnissen liefern. Dies ist ein erster Schritt zur Definition von standortspezifischen Referenzwerten. Referenzwerte wiederum können eine Grundlage für die Entwicklung von Bewertungsinstrumenten für eine angewandte Nutzung durch Politik, Bodenmanagement und Regulierungsbehörden bilden.

Ein weiteres Ziel dieser Studie ist, Muster der Artenverbreitung und Häufigkeitsbeziehungen aufzudecken, die zur Erhaltung der Regenwurmarten in Deutschland beitragen können.

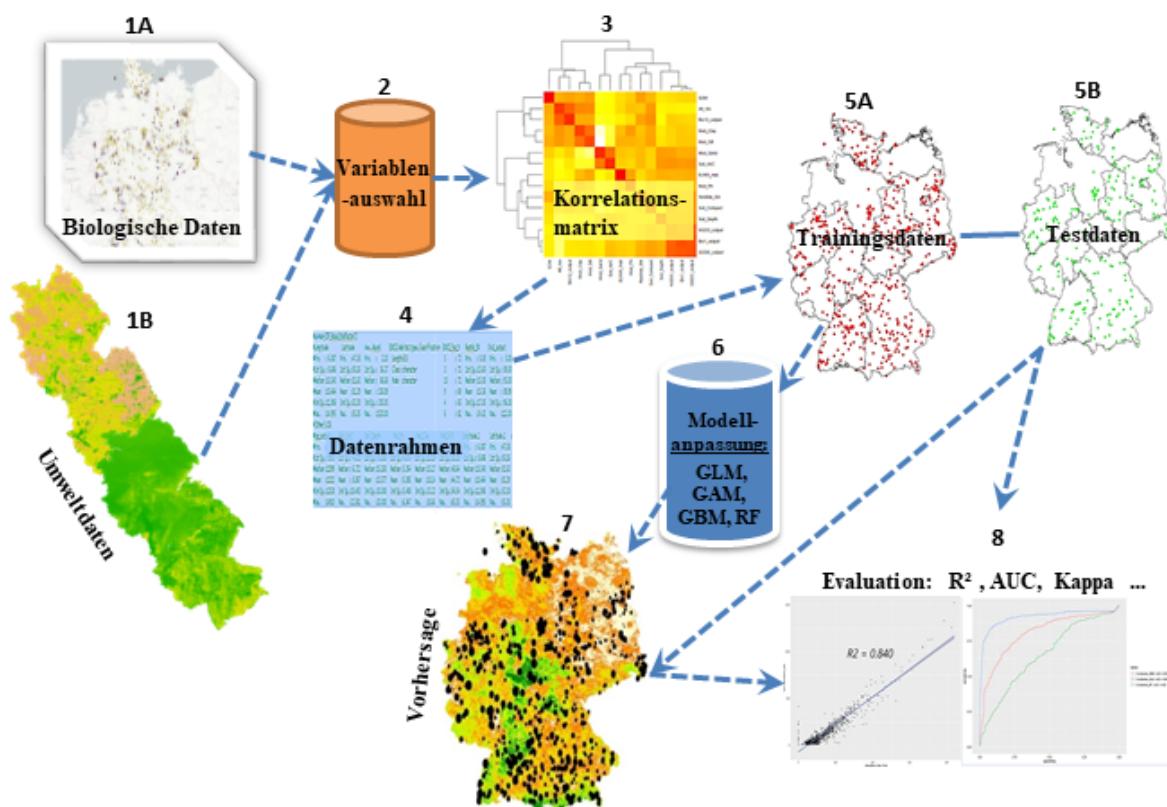
2 Material und Methoden

2.1 Räumlicher Umfang und Workflow

Der räumliche Umfang der Studie umfasst den gesamten Bereich innerhalb der politischen Grenzen Deutschlands. In den Modellen und für die Modellalgorithmen wurde der Breitengradbereich von $47,3209 - 54,9049^\circ$ N und der Längengradbereich von $6,0470 - 14,8428^\circ$ E festgelegt (s. Abbildung 2).

Der Modellierungsworkflow des Projektes (Abbildung 1) besteht aus acht systematischen Schritten: Zusammenstellung von (1A) biologischen und (1B) Umweltdaten, (2) Variablenauswahl und Modellkalibrierung, (3) Multikollinearitätstest, (4) Vereinheitlichung der Eingangsdatenformate, (5) Aufteilung des Datensatzes in Trainings- und Testdaten, (6) Modellkalibrierung der Regenwurmhabitatemperatur, (7) Vorhersage (Upscaling der Modellergebnisse zu Karten) und (8) Modellauswertung und Beurteilung.

Abbildung 1: Übersicht über den Arbeitsablauf der Studie



Quelle: Eigene Darstellung, BGR und Senckenberg Museum

2.1.1 Zusammenstellung der Regenwurmdaten (Arbeitsschritt 1A)

2.1.1.1 Datenabfrage bei den Staatlichen Geologischen Diensten

Es wurde geprüft, ob es in den Dauerbeobachtungstellen der Bundesländer bodenbiologische Untersuchungen gibt, deren Ergebnisse noch nicht in den Edaphobase⁹-Datenbestand übernommen wurden. Rückmeldungen kamen aus sieben von 16 Bundesländern (Brandenburg, Baden-

⁹ <https://portal.edaphobase.org>

Württemberg, Bayern, Hessen, Schleswig-Holstein, Saarland und Sachsen-Anhalt). Die Daten aus Brandenburg, Baden-Württemberg, Sachsen-Anhalt und Schleswig-Holstein waren bereits vollständig in Edaphobase aufgenommen. Aus Bayern wurden einzelne Funddaten nachgetragen. In Hessen und dem Saarland stehen keine bodenbiologischen Daten zur Verfügung.

2.1.1.2 Datenharmonisierung und -bereinigung

Die Regenwurmrohdaten wurden am 22. Februar 2021 aus Edaphobase unter Verwendung der Filter „Lumbricidae“ & „Deutschland“ heruntergeladen (Abbildung 2). Das Ergebnis war ein Datensatz bestehend aus 22 134 Einzeldatenaufzeichnungen (Datensätze). Alle Metadaten bezüglich Boden, Habitattypen, Klima usw., die in Edaphobase zu Orten mit Regenwurmvorkommen verfügbar waren, wurden heruntergeladen und mit den Funddatensätzen verbunden. Zwei Arten von Daten waren verfügbar: Daten zu Vorkommen (einfache Präsenz der Art, registriert bei Vorhandensein am jeweiligen Ort) und/oder Daten zur Abundanz (Anzahl der Regenwürmer je Art am jeweiligen Ort). Nur Abundanz-Daten harmonisiert auf Individuen pro Quadratmeter (Ind. m⁻²) wurden zur Berechnung der Gesamtabundanz verwendet (Gesamtabundanz der Regenwurmgesellschaft oder Populationsgröße einer individuellen Art), da Anzahlen (counts), (d. h. die pure Anzahl von einzelnen Regenwürmern an einem Ort und/oder Datum) nicht zu vergleichen sind aufgrund möglicherweise verschiedener Probenflächengrößen.

Der Edaphobase-Rohdatensatz war heterogen, da er aus verschiedenen Quellen und von verschiedenen Datenbereitstellern stammt. Die meisten Daten gab es auf der Gemeinschaftsebene (d. h. alle bei der Probennahme an einer Stelle präsenten Arten). Andere Daten betrafen nur Beobachtungen zu einzelnen Arten. Während weiterhin einige Daten Durchschnittswerte für einen einzelnen Standort angaben, waren andere Daten differenziert hinsichtlich des Datums der Probennahme an einem Ort (Werte zum Beobachtungsdatum); einige Daten bezogen sich sogar auf individuelle Proben. Zusätzlich waren einige Daten von verschiedenen Probengebieten in einem Geokoordinatensatz zusammengefasst, so dass verschiedene Probennahmenstellen (*plots*) fälschlicherweise von einer Stelle zu stammen schienen. Diese Heterogenität erforderte ein intensives Bearbeiten des Edaphobase-Rohdatensatzes, bevor er für das Modellierungsverfahren genutzt werden konnte. Der Rohdatensatz wurde daher zunächst harmonisiert und bereinigt, um alle Daten in eine vergleichbare Struktur und Formatierung zu bringen, so dass ein allgemeingültiger Einsatz für die Modelle möglich war. Um kohärente ortsspezifische Datensätze zu erhalten, wurden die Daten gefiltert nach: (1) Breitengrad, (2) Längengrad, (3) Datum der Beobachtung und (4) gültigem Artnamen. Daten ohne Geokoordinaten wurden aus dem Datensatz gelöscht.

Zuerst wurde der Rohdatensatz auf Gemeinschaftsebene bereinigt (Daten zu dem gesamten vor kommenden Artenbestand an einem individuellen Ort, der während einer Studie bzw. Untersuchung festgestellt wurde). Nur Daten mit gleichen Geokoordinaten und aus mindestens zwei Arten bestehend wurden als Daten auf Gemeinschaftsebene gewertet und aufgenommen. Datenaufzeichnungen von Beobachtungen nur einer Art wurden nur in den wenigen Fällen im Datensatz behalten, wenn aus den Metadaten für Quelle, Habitattyp oder Bemerkungen plausibel hervorging, dass es sich um Beobachtungen zur Gemeinschaftsebene handelte. Alle in dieser Hinsicht zweifelhaften Datenaufzeichnungen von Beobachtungen einzelner Arten wurden aus dem Datensatz gelöscht.

Abbildung 2: Räumliche Verteilung der Regenwurmfunddaten



Quelle: Eigene Darstellung, Edaphobase, BGR

Daten aus Höhlen wurden gelöscht, ebenso Daten von urbanen Standorten, wie z. B. Pflanzcontainern in der Stadt oder auch Verkehrsinseln, da (1) solche Daten auf wenige Bereiche begrenzt sind und (2) die Modellierungsverfahren auf Regenwürmer in (natürlichen) Böden abzielen. Darüber hinaus wurden auch sehr alte Daten gelöscht, die artenarme Gemeinschaften betreffen

(ca. vor 1920, meist ältere Museums-Aufzeichnungen), da diese höchst wahrscheinlich nicht Daten auf Gemeinschaftsebene widerspiegeln.

Insgesamt resultierte daraus ein Datensatz auf Gemeinschaftsebene aus 21 395 Einzeldatensätzen (inklusive Einzel-Artenaufzeichnungen). Der Großteil der ursprünglich abgefragten Edaphobase-Daten (96,7 %) ist auf der Gemeinschaftsebene nutzbar.

Im Anschluss daran wurden die Daten eines Ortes nach Datumsangaben aufgeteilt (sofern diese existierten). Einige Aufzeichnungen beinhalteten kein Beobachtungsdatum. Für den Fall, dass diese Daten nicht plausibel mit einer Regenwurmgemeinschaft kombiniert werden konnten (eines Ortes und eines Beobachtungsdatums), wurden sie gelöscht. Somit wurde sichergestellt, dass nur korrekte Daten auf Gemeinschaftsebene für die Modellierung genutzt wurden. Für Standorte, die über einen längeren Zeitraum beobachtet wurden (z. B. Bodendauerbeobachtungsflächen, Sukzessionsstudien o. ä.), wurden nur die jüngsten Beobachtungen (der letzten 10 Jahre) berücksichtigt, da sie am wahrscheinlichsten die momentanen Regenwurmvorkommen repräsentieren. Ganz ähnlich wurden Daten sehr früher Sukzessionsstadien (z. B. in Bergbaufolgelandschaften) gelöscht und nur die jüngsten Sukzessionsstadien behalten. Wenn in dem verbleibenden Datensatz Daten desselben Standortes existierten, die zwei verschiedene Habittypen aufgrund eines unterschiedlichen Datums der Probennahme aufwiesen (durch Änderungen in der Landnutzung oder Habitatsukzession), wurden diese wie zwei verschiedene Standorte behandelt (es handelte sich dabei meistens um Ackerland, das in Grünland umgewandelt wurde bzw. umgekehrt). In den wenigen Fällen, in denen Daten für denselben Ort aus verschiedenen (fast immer zwei) Quellen vorlagen und diese extrem variierten (z. B. in der Artenanzahl), wurde die plausiblere oder aus vertrauenswürdigerer Quelle stammende Angabe verwendet und die andere gelöscht. All diese Maßnahmen reduzierten den finalen Datensatz auf 20 013 individuelle Datensätze, der somit 91,4 % der ursprünglichen Edaphobase-Rohdaten repräsentiert.

2.1.1.3 Parameter auf Gemeinschaftsebene (Gesamtabundanz und Artenreichtum)

Der resultierende finale Rohdatensatz wurde dann zu einzelnen Gemeinschaftsebene-Werten je Standort zusammengefasst für den Einsatz in den Modellen. Für das Modellieren von Gemeinschaften wurden alle artenspezifischen Daten eines Standortes zusammengefasst. So erhielt man die Parameter Regenwurm-Gesamtabundanz (harmonisiert auf die Anzahl der Regenwurm-Individuen pro Quadratmeter) sowie Regenwurm-Artenreichtum (durchschnittliche Anzahl der Arten, die an einem Standort gefunden wurden). Oft gibt es nur Daten zur Anzahl der einzelnen Regenwürmer einer Art, die an einer Stelle und an einem Datum identifiziert wurden. Da solche Daten keinen räumlichen Bezug angeben, repräsentieren sie nicht die wirkliche Vielfalt (= Dichten) und sind nicht vergleichbar mit den anderen Daten zur Vielfalt (harmonisiert auf Ind. m⁻²). Sie flossen daher nicht in die Zusammenfassung zur Abundanz ein, wurden jedoch für die Aggregation zu Artenreichtumswerten verwendet, da sie das Artenvorkommen an einer Stelle angeben. Hinsichtlich einiger sehr großer Datensätze aus bestimmten individuellen Quellen (z. B.

Burkhardt¹⁰, Luthardt¹¹, Neubert¹² oder Walter¹³) wurden die Probengrößen durch persönliche Kontaktaufnahme oder Literatur bestimmt und die Zählungen dann übertragen auf Ind. m⁻² (Abundanz). Einige Abundanzdaten wurden offensichtlich falsch in Edaphobase eingegeben und repräsentierten nur Anzahlen (z. B. bestanden die Angaben nur aus ganzen Zahlen, hatten sehr niedrige Werte und enthielten oftmals den Wert 1); diese Datensätze wurden nicht für die Abundanzberechnungen genutzt.

Für die Berechnung der Gemeinschaftsgesamtabundanz (*total community abundance*) eines Standortes wurde die durchschnittliche Vielfalt jeder Art für jedes Datum der Probennahme summiert (oder für den Ort, falls Daten nur für ein oder kein Datum vorlagen). Wenn im Vorfeld Daten für eine spezielle Stelle verschiedene Aufzeichnungen von individuellen Arten enthielten (Daten für den Punkt Probennahme), wurde die Vielfalt jeder Art im Durchschnitt für jeden Standort und jedes Datum angenommen (wenn nur ein Datum, dann für den Ort). Wenn Daten für einen bestimmten Ort mit mehreren Datumsangaben vorlagen, wurde der Vielfaltsdurchschnittswert berechnet; dieser wurde dann als der Gesamtwert für die Gemeinschaftsvielfalt (*total community abundance value*) des Standortes angenommen. Wenn kein Datum zur Probennahme („*no data*“) in den Datenaufzeichnungen für einen bestimmten Ort angegeben wurde, wurde dies wie ein einziges Datum zur Probennahme behandelt und die durchschnittliche Vielfalt für jede Art berechnet.

Für den Parameter Artenreichtum einer Gemeinschaft (*community species richness*) wurde die Anzahl der verschiedenen Arten, die an einem Ort zu einem Datum aufgefunden wurden, summiert. Daten, die nur für den Punkt Gattung vorlagen, wurden nur als separate Art gewertet, wenn keine andere Art der Gattung für den Ort und das Datum angegeben war (alle Daten im Feld Familie [*Lumbricidae*] flossen nicht in die Artenreichtumsberechnungen mit ein, aber die Vielfaltsdaten für den Punkt Familie wurden gespeichert für die Berechnungen der gesamten Gemeinschaftsvielfalt, s. o.). Wenn Daten zu einem unterschiedlichen Datum an einem Ort vorlagen, wurde – wie bei den Vielfaltsdaten – der Durchschnittswert der Artenreichtumswerte eines jeden Datums genommen, um einen Wert für jeden Ort zu erhalten. Es wurde entschieden, die Durchschnittswerte der Anzahl der aufgezeichneten Arten eines jeden Datums zu nehmen, anstatt der Gesamtzahl der aufgezeichneten Arten an einem Ort, da sich die Daten zur Probennahme oft über einen Zeitraum von mehreren Jahren erstreckten und nicht sichergestellt werden konnte, dass vor vielen Jahren aufgezeichnete Arten immer noch an diesen Orten vorkamen.

Um Daten zu entfernen, die möglicherweise außerhalb des geographischen Bereichs des Studiengebietes lagen, wurde der Punkt-in-Polygon-Filter unter Einsatz der `point.in.polygon` Funktion des „sp“ package (Pebesma & Bivand, 2005) in R benutzt (R core team, 2020), wobei der

¹⁰ Burkhardt, U. et al. (2010). Unveröffentlichter Sammlungsbestand Lumbricida 1960 – 2008, SMN Görlitz [Datensatz].

¹¹ Luthardt, V. et al. (2009). Lebensräume im Wandel: Ergebnisse der ökosystemaren Umweltbeobachtung im Biosphärenreservat Spreewald. Fachbeiträge des LUGV. 122, 1-158.

Luthardt, V. & Hagenmüller, K. (2012). Projekt Ökosystemare Umweltbeobachtung ÖUB [Datensatz].

¹² Neubert, E. (2012). Rohdaten Regenwürmer Landesamt für Umweltschutz Sachsen-Anhalt 2004 – 2010 [Datensatz].
Neubert, E. (2012). Lumbricidenuntersuchungen auf ausgewählten Boden-Dauerbeobachtungsflächen (BDF) in Sachsen-Anhalt. Ergebnisbericht 2011.

Neubert, E. (2014). Lumbriciden 2013. Regenwürmer auf ausgewählten Boden-Dauerbeobachtungsflächen (BDF) in Sachsen-Anhalt.

¹³ Walter, R. (2013). Bayrische Bodendauerbeobachtungsflächen [Datensatz].

punktuelle Datensatz zu einem Vorkommen ausgeschnitten und mit dem geographischen Bereich des Studiengebietes abgeglichen wurde. Für diese Funktion wurde ein „*boundary shapefile*“ von Deutschland aus de GisDataCollection (igismap.com) eingesetzt.

Diese Verfahren zur Zusammenführung resultierten dann in einem finalen Artenreichtums-Datensatz auf Gemeinschaftsebene (*species-richness, community-level*) von 992 Standorten. Da Daten von vielen Standorten keine Informationen zur Vielfalt aufwiesen (nur die Präsenz), wurden diese nicht in den Datensatz zur Vielfalt aufgenommen (aber für den Datensatz Artenreichtum gespeichert). Dies reduzierte den Datensatz zum Punkt gesamte Vielfalt der Gemeinschaft „*community-level total abundance*“ auf 653 Standorte in ganz Deutschland.

2.1.2 Datensatz auf Arrebene

Um die potentielle Verbreitung von individuellen Arten zu modellieren – das o. g. Datenaufbereitungsverfahren war extrem zeitaufwändig – wurde ein „*Species-level*“-Datensatz vom ursprünglichen aufbereiteten Gemeinschaft-Datensatz abgeleitet (von 21.395 Funddatensätzen). Einige Einzelarten-Funde aus Edaphobase (z. B. jene, die beim o. g. Aufbereitungsverfahren gelöscht wurden), sind daher nicht verwendet worden. Dies ist aus den folgenden Gründen zu rechtfertigen: (1) weniger als 2 % der gesamten Edaphobase-*Lumbricidae*-Daten wurden während der ersten routinemäßigen Datenaufbereitung gelöscht und (2) betrafen diese gelöschten Daten Arten, die in Deutschland wenig verbreitet sind und zu wenig Daten zu ihrer Verbreitung vorliegen, so dass eine Modellierung nicht möglich gewesen wäre.

Bezüglich der artenspezifischen Daten (innerhalb einer Gemeinschaft) wurden die Daten für jede Art mit den Parametern Vorkommen (*occurrence*) (ja/nein [=1/0] für jede Art und jeden Ort) und Artenabundanz (wieder harmonisiert auf einzelne Funde m^{-2}) für jeden Ort zusammengeführt. Wir bearbeiteten also die Artenlisten hinsichtlich der Vielfalt sowie für Präsenz (1) und Absenz (0) s. Anlage. Die gleichen Verfahren wie oben wurden zum Zusammenführen der Daten von individuellen Arten genutzt: wenn Daten von verschiedenen Proben für einen Standort vorlagen, wurde der Durchschnitt der Artenvielfalt für jedes Datum genommen. Die berechneten durchschnittlichen Werte dienten zur Bestimmung der durchschnittlichen örtlichen standortsbezogenen Artenvielfalt. Wenn nur ein (oder kein) Datum zur Probennahme für einen Standort angegeben war, aber verschiedene Proben vorlagen, wurde die gesamte (durchschnittliche) Vielfalt für jede registrierte Art berechnet. Wenn nur eine Aufzeichnung für den Standort vorlag, wurde der vorliegende Wert für die Artenvielfalt des Ortes genutzt. Hinsichtlich der Artenvorkommen wurden individuelle Arten als vorkommend [1] an einem Ort gelistet, wenn Daten für diese Art an einem beliebigen Datum an dieser Stelle vorlagen. Da Daten auf Arrebene von den Rohdaten auf Gesellschaftsebene abgeleitet wurden und während des Aufbereitungsverfahrens als aussagekräftig für die Studien auf Gemeinschaftsebene eingestuft wurden, konnte angenommen werden, dass jene Arten, die nicht in den Daten für diesen Ort aufgelistet waren, nicht oder nur bedingt an dieser Stelle vorkamen (jedenfalls nicht an dem Zeitpunkt der Probennahme(n)). In solchen Fällen wurden diese Arten als nicht vorkommend an dieser Stelle aufgelistet (= Pseudo-Absenz [0]). Das Ergebnis war ein Datensatz auf Arrebene von 45 Arten an insgesamt 992 Orten.

Einseitige Probennahme und Datenautokorrelation aufgrund von Datenclustern sind bei den meisten Datensätzen zum Vorkommen, die für SDM zusammengestellt wurden, zu finden. Dies kann zu einer umwelt- und vorhersagetechnischen Einseitigkeit bei der Biodiversitätsmodellie-

rung führen (Boria *et al.*, 2014; Aiello-Lammens *et al.* 2015). Um diese Thematik bei der Modellierung zur Vorhersage und Kartierung der Verbreitung und geographischen Ausbreitung der einzelnen Regenwurmarten in Deutschland zu verbessern, haben wir eine räumliche Ausdünnung („*spatial thinning*“) bei den Aufzeichnungen der Arten (Vorkommen) vorgenommen, und zwar unter Einsatz der „*thin*“ Funktion der R „*SpThin*“ package (Aiello-Lammens *et al.* 2015). Räumliche Ausdünnung wurde ebenso zur Problemlösung von Datenautokorrelation bei der ökologischen Modellierung vorgeschlagen, da Autokorrelation dazu neigt, bei Distanzen, die größer als der gegebene Wert sind, geringer zu werden (Fortin & Dale, 2005). Wir haben eine Minimum-Schwellendistanz von 1 pro Zelle im Rasterfile (Fourcade *et al.*, 2014) angesetzt, um einen zu hohen Datenverlust bei hohen Artenvorkommen zu vermeiden.

Der finale Datensatz enthielt somit 45 gültige Arten. Während alle Arten zur Modellierung der gesamten Gemeinschaftsvielfalt und des Artenreichtums genutzt wurden, haben wir für diese Studie – aufgrund der eingeschränkten Datenverfügbarkeit beim Großteil der Arten – nur 12 Arten zur Modellierung und Kartierung ihrer räumlichen Verbreitung ausgewählt. Unsere Auswahl begründete sich auf (1) die am meisten verbreiteten Arten basierend auf der Menge der verfügbaren Arten, auf Literatur und Expertenwissen; (2) Arten mit einer offensichtlich einzigartigen Habitatpräferenz; (3) Arten, die als selten angenommen wurden; und (4) Arten, von denen angenommen wurde, dass sie endemisch sind (Tabelle 1).

Tabelle 1: Für die Modellierung der potentiellen Artenverbreitung ausgewählte Regenwurmarten

Art	Anzahl Fundorte	Ökologische Lebensformengruppe	Auswahlkriterien
<i>Aporrectodea caliginosa</i> SAVIGNY, 1826	719	endogäisch	gewöhnlich
<i>Aporrectodea rosea</i> SAVIGNY, 1826	603	endogäisch	gewöhnlich
<i>Lumbricus terrestris</i> LINNAEUS, 1758	546	anözisch	gewöhnlich
<i>Lumbricus rubellus</i> HOFFMEISTER, 1843	595	epigäisch	gewöhnlich
<i>Allolobophora chlorotica</i> SAVIGNY, 1826	304	endogäisch	gewöhnlich
<i>Lumbricus castaneus</i> SAVIGNY, 1826	319	epigäisch	gewöhnlich
<i>Aporrectodea longa</i> UDE, 1885	143	anözisch	selten?
<i>Dendrobaena octaedra</i> SAVIGNY, 1826	309	epigäisch	besonderes Habitat
<i>Dendrobaena attemsi</i> MICHAELSEN, 1903	29	epigäisch	selten?
<i>Allolobophoridella eiseni</i> LEVINSEN, 1884	43	epigäisch	selten?
<i>Aporrectodea limicola</i> MICHAELSEN, 1890	67	endogäisch	besonderes Habitat

2.1.3 Umweltdaten (Arbeitsschritt 1 B)

2.1.3.1 Zusammenstellung von Umweltdaten

Mit höchster Priorität wurden die in Edaphobase bei den individuellen Regenwurmfundpunkten abgelegten Umweltinformationen für die Kalibrierung der Modelle verwendet. Fehlende Umweltinformationen wurden aus den folgenden Quellen in der Datenbank oder im Zuge der Modellkalibrierung aus den Rasterdatensätzen ergänzt.

Klimavariablen wurden von *Climatologies at High resolution for the Earth Land Surface Areas* (CHELSEA, s3_browsind, wsl.ch) bezogen. Dies waren im Besonderen: mittlere Jahrestemperatur („Bio1“), die gesamten jährlichen Regenfälle („Bio12“), die durchschnittliche Vegetationsperiode („growing degree days“) mit einer Temperatur über 10°C („NGD10“) und die durchschnittliche jährliche Anzahl von frostfreien Tagen mit einer Temperatur über 0°C („GDD0“). Diese Klimadaten enthalten Durchschnittswerte der Jahre 1981 bis 2010 mit einer räumlichen Auflösung von 30 Bogensekunden, was dem Rahmen der räumlichen Auflösung der biologischen Daten entspricht.

Für die topographische Variablen (Meereshöhe) wurde der USGS-NASA-Earthexplorer-Datensatz¹⁴ verwendet.

Zur Ergänzung der Bodenvariablen (für 571 Fundpunkte) wurden folgende Datenquellen geprüft:

- ▶ Geologische und bodenkundliche Daten mittleren Maßstabs der Staatlichen Geologischen Dienste der Länder
- ▶ Daten des European Soil Data Centre (ESDAC):
 - LUCAS TOPSOIL-Daten 2009 und 2015
 - LUCAS land use/land cover 2015
 - europaweite Karten der chemischen Bodeneigenschaften auf der Basis der LUCAS-Oberbodendaten 2009/2012¹⁵ (pH der Bodenlösung in Wasser sowie CaCl₂, C/N-Verhältnis, Kationenaustauschkapazität, Calciumcarbonat-, Stickstoff-, Phosphor- und Kaliumgehalt) (Ballabio *et al.*, 2016, 2019)
- ▶ Bodenflächendaten der Bundesanstalt für Geowissenschaften und Rohstoffe im Maßstab 1 : 1 Mio. bzw. 1 : 200 000 (Sandgehalt, Schluffgehalt, Tongehalt, Gründigkeit, effektive Lagerungsdichte, Luftkapazität [Porosität] und Gehalt an organischer Substanz im Boden [SOM]). Als Rasterdatensätze mit einer räumlichen Auflösung von 250 m x 250 m wurden diese Daten auch in der Regionalisierung durch die SDM verwendet.
- ▶ European Space Agency (ESA), Climate Change Initiative (CCI)¹⁶ (durchschnittliche jährliche Bodenfeuchte als Bodenwassergehalt in 2,5 cm Bodentiefe)

¹⁴ <https://earthexplorer.usgs.gov/dataset>

¹⁵ <https://esdac.jrc.ec.europa.eu/content/chemical-properties-european-scale-based-lucas-topsoil-data>

¹⁶ <https://www.esa-soilmoisture-cci.org/data>

Wegen der bundesweiten Einheitlichkeit der Daten und wegen ihrer geeigneten Auflösung im Vergleich zu den biologischen Daten sind letzten Endes vor allem die Bodendaten der BGR, insbesondere der Bodenübersichtskarte 1 : 200 000 (BÜK 200), in die Modelle eingeflossen. Weitere Bodenvariablen sind bekanntermaßen von hoher Relevanz für das Vorkommen von Regenwürmern und Regenwurmpopulationsgrößen, wie der Stickstoffgehalt des Bodens und das C/N-Verhältnis im Boden als Indikator für die Qualität der organischen Bodensubstanz. Diese lagen jedoch nicht bzw. nicht in ausreichender Auflösung vor und konnten daher nicht in den Modellierungsverfahren verwendet werden. Auch mikrobielle Parameter dürften von hoher Relevanz sein, allerdings liegen auch hierzu keine deutschlandweiten Daten vor.

Räumlich höher aufgelöste Daten der Staatlichen Geologischen Dienste der Länder wurden ausführlich auf ihre Eignung geprüft. Letztlich sprachen die hohe Heterogenität der verfügbaren Daten und zum Teil nicht passfähigen Datenstrukturen gegen eine Verwendung in diesem Projekt.

Die in Edaphobase vorliegenden Fundortnachweise der Gruppe der Regenwürmer (Lumbicidae) ohne ökologische Begleitparameter wurden anhand von Daten der BÜK 200 ergänzt. Unter anderem wurden Daten zur Habitatpräferenz bzw. die Einordnung in funktionelle Gruppen (epigäisch, endogäisch, anözisch), Bodenart nach Feinbodenartendiagramm (7646 von 11946), Humusgehaltklassen in Prozent-Spannen (10657 von 11946), Bodenausgangsgestein (11900 von 11946), Durchlässigkeit (11855 von 11946) und die Verknüpfung mit den Legendeneinheiten der BÜK 200 (alle Datensätze) ergänzt. Über die Verknüpfung mit den Legendeneinheiten der BÜK 200 lassen sich Fundortdaten mit zusätzlichen Parametern wie Humusgehalt oder pH-Werten verknüpfen.

Landnutzungs-/Landbedeckungsdaten wurden aus der EUNIS-Klassifizierung europäischer Ökosystemtypen bezogen, basierend auf Details der Habitatklassifizierung, Stufe 1, Version 3.1, 2019 mit einer räumlichen Auflösung von 100 m¹⁷. Die Habitat-Rasterdaten wurden in 11 Landnutzungsklassen gruppiert (meeresbeeinflusst, Küstengewässer, Binnengewässer, Moore und Moore, Grünland, Heide, Wald, spärliche Vegetation, Acker, urbane und künstliche Habitate und nicht klassifizierte Habitatkomplexe; Tabelle 3). Jeder der Klassen wurde eine Proxy-Nummer von 1 – 10 zugeordnet, wobei 0 für nicht klassifizierte oder komplexe Lebensraumtypen steht.

2.1.3.2 Auswahl der ökologischen Variablen

Diejenigen Umweltdaten, die in den Regenwurmfunddaten begleitend aufgenommen wurden, wurden bei der Kalibrierung der Algorithmen als Schätzwert für Regenwurmvorkommen genutzt. Die Auswahl geeigneter Schätzwertvariablen stellt einen wichtigen Schritt bei der SDM dar, da die Ausprägung der Umweltparameter als Beschränkung der Artenverbreitung wirkt und die ökophysiologischen Toleranzen und Bedürfnisse der Arten widerspiegeln. Daher müssen die ökologischen und physiologischen Variablen hinsichtlich ihrer Relevanz sorgfältig ausgewählt werden (Guisan & Zimmermann, 2000; Austin, 2002; Jarnevich *et al.*, 2015). Das Hauptkriterium zur Auswahl der Schätzwertvariablen war, dass sie einen physiologischen Einfluss auf die Verbreitung der Regenwurmarten haben und eine ökologische Bedeutung tragen (Guisan & Zimmermann, 2000; Rutgers *et al.*, 2016; Phillips *et al.*, 2019; Edwards und Arancon, 2022). Die im Vorfeld analysierten und ausgewählten Variablen wurden in vier breite Kategorien eingruppiert: Klima- (4 Variablen), topographische (1), Boden- (9) und Habitat-/Landnutzungs-Klassen.

Von allen Umweltdaten haben wir anfänglich 15 externe Schätzwertvariablen ausgewählt, die auf relevanten umweltspezifischen Parametern basieren und im Zusammenhang mit Artenvor-

¹⁷ <https://www.eea.europa.eu/data-and-maps/data/ecosystem-types-of-europe-1>

kommen in Edaphobase, Literatursuche, Online-Recherchen und Experten-basierten Einschätzungen in Verbindung stehen. Alle Umweltdaten, als solche bzw. abgeleitete Werte, wurden als räumlich explizite Rasterdatensätze für die räumlichen Modellierungen verwendet (Tabelle 2).

Da die in Edaphobase enthaltenen Umweltdaten spezifisch und direkt mit den Regenwurmvorkommen auf Fundpunktebene verknüpft sind und in der Regel gleichzeitig mit den Regenwurmdaten erhoben wurden, wird davon ausgegangen, dass die Umweltdaten in Edaphobase einen wesentlich genauer und damit für die Modellierung der Artansprüche geeigneter sind als die externen Bodendaten oder andere der oben aufgeführten Umweltdaten. Obwohl die Fundpunkte

Tabelle 2: Liste vorausgewählter Umweltparameter, die als Prädiktorvariablen verwendet werden (Datenquellen sind im Text angegeben)

Code	Beschreibung	Räumliche Auflösung	Einheit
Bio 1	Jahresmitteltemperatur	1 km	°C (Beachte: im Datensatz °C*10)
Bio 12	Mittlerer Jahresniederschlag	1 km	mm
Mod_clay	Tongehalt des Bodens	250 m	Prozent (%)
Mod_silt	Schluffgehalt des Bodens	250 m	Prozent (%)
Mod_sand	Sandgehalt des Bodens	250 m	Prozent (%)
SoilAirC	Luftkapazität des Bodens	250 m	Volumen- Prozent (Vol.-%)
SOM	Gehalt an organischer Substanz des Bodens	250 m	Prozent (%)
ResMsk_SM/S_Moisture	Bodenfeuchte	250 m	% Bodenwasser-gehalt in 2 – 5 cm Tiefe
Soil_compact	Lagerungsdichte des Bodens	250 m	Trockenrohdichte g cm ⁻³
Soil_depth	Gründigkeit des Bodens	250 m	Tiefe in cm
Mod_pH	pH (0,01 M CaCl ₂)	250 m	dimensionslos
EUNIS Hab	Habitat-/Landnutzungstyp	100 m	nominalskaliert (s. Tabelle 3)
Alt_res	Meereshöhe	90 m	Meter
NGDO_output	Dauer der Vegetationsperiode (Tage mit mittlerer Lufttemperatur über 10 °C)	1 km	Durchschnittliche Zahl von Tagen pro Jahr
GDDO_output	Zahl der Tage mit Lufttemperaturen über 0 °C	1 km	Durchschnittliche Zahl von Tagen pro Jahr

aus Edaphobase eine relativ große Fläche abdecken, gibt es große nicht beprobte Gebiete, so dass in der Fläche insgesamt eine hohe Datendichte vorliegt, aber mit großen, nicht abgedeckten Gebieten. Deshalb wurden in den für die Modellkalibrierung verwendeten Rasterdateien die Zellen mit Edaphobase-Daten mit den Umwelt-Daten aus Edaphobase belegt. Insbesondere für den

pH-Wert des Bodens wurden dabei die fehlenden Werte in den Punktdaten vorhergesagt aus den Rasterdaten, die mit den Bodentexturparametern (Ton-, Sand- und Schluffgehalt) als Kovariaten kombiniert wurden. Die Auswahl der zu ersetzenen Zellen der Rasterdatei erfolgte auf der Grundlage der Geokoordinaten der Vorkommensorte. Nach der Modellkalibrierung basieren die Vorhersagen und Kartierungsalgorithmen auf den ursprünglichen (externen) Rasterdaten (ohne Ersatz durch Edaphobase-Daten).

In einigen Zellen der Rasterdaten waren keine Daten verfügbar (NA), z. B. für die Bodentrockenrohdichte, Luftkapazität und Gründigkeit. Die fehlenden Daten in den Rasterdateien dieser Vari-

Tabelle 3: Europäische EUNIS-Klassifikationsebenen für Habitat-/Landnutzungstypen auf Ebene 1, die als Modellierungskategorien der Habitattypvariablen verwendet werden (EUNIS Hab, Tabelle 2)

Code	Habitat-/Landnutzungstyp
LULC_1	Meeresbeeinflusste Habitate (d. h. Meeresinseln)
LULC_2	Küstenhabitale
LULC_3	Binnenwasserkörper
LULC_4	Hoch- und Niedermoore
LULC_5	Grünland sowie kraut-, moos- oder flechtendominierte Habitate
LULC_6	Heiden, Gebüsche und Tundra
LULC_7	Wald, Forst
LULC_8	Ödland
LULC_9	Ackerland
LULC_10	Siedlungen, Industrie- und Gewerbegebiete, künstliche Habitate
0	nicht klassifiziert/komplex

ablen erzeugen Lücken, die die endgültigen modellierten Vorhersagen und Karten verzerren könnten. Um dies zu beheben, wurden die Lücken mit dem „Gap Fill“-Tool des System for Automated Geoscientific Analyses (SAGA) GIS geschlossen. Dies erfolgt durch Zuweisen von in den Eingabe-Rasterkarten verfügbaren Datenwerten zu den NA-Zellen in den Regionen mit Datenlücken. Die Werte für die NA-Datenregionen wurden aus den umgebenden Zellwerten mit inverser Distanzgewichtung berechnet. Nach dieser Interpolation wurden die Ergebnisse geglättet (GDAL-Software-Suite, 2013).

Für den pH-Wert des Bodens enthielten die verfügbaren Rasterdateien für Deutschland große räumliche Bereiche mit fehlenden Daten. Um die Genauigkeit der pH-Daten als Umweltprädiktorvariablen zu verbessern, wurde ein Modellierungsverfahren zum Füllen von Datenlücken verwendet (Hengl, 2007; Reuter *et al.*, 2008; Soil Security Lab 2017; Wieger Wamelink, 2019). Edaphobase enthielt viele pH-Punktdaten innerhalb dieser Bereiche mit Datenlücken. Da diese Felddaten relativ ausreichend waren, haben wir die Beziehung zwischen pH-Felddaten und ihren Kovariaten (d. h. Bodentexturdaten) an beprobten Standorten modelliert und den pH-Wert für nicht beprobte Gebiete basierend auf dieser Beziehung vorhergesagt, was zu einem modifizierten Umweltraster für den pH-Wert führte.

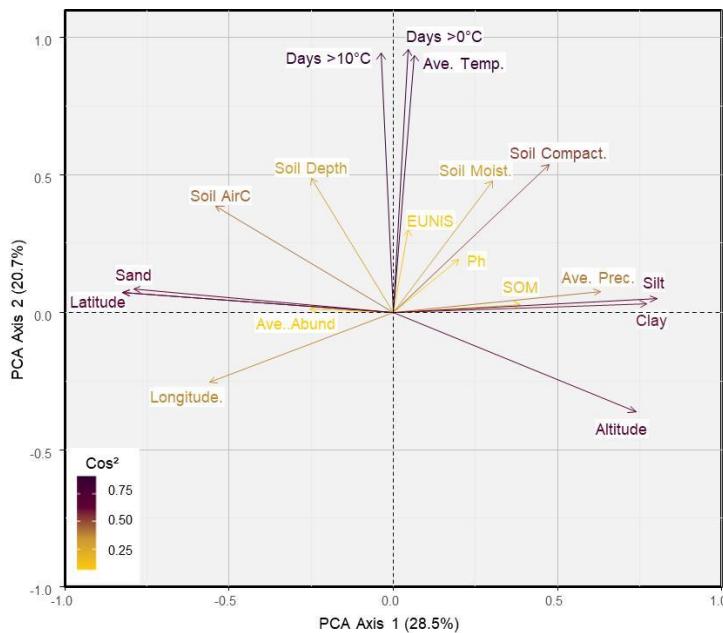
Da alle Umwelt-Rasterdaten aus externen Quellen mit unterschiedlichen Formaten, Auflösungen, Ausdehnungen, usw. bezogen wurden und die SDM-Modellierung erfordert, dass diese Variablen in geographischer/räumlicher Ausdehnung einheitlich sind, wurden zunächst alle vorausgewählten Variablen auf das Weltweite Geographische Koordinatensystem WGS84 umreferenziert, in dem alle Edaphobase-Daten referenziert sind, beschnitten und in das GIS-Shapefile für Deutschland maskiert und auf eine einheitliche räumliche Auflösung von 250 m gesampled bzw. disaggregiert, um der Auflösung der Punktdaten der Arten zu entsprechen, wobei das 3. Quartil der räumlichen Auflösung 260 m betrug (Thuiller *et al.*, 2004; Guisan *et al.*, 2017). Diese prozessierten Umweltvariablen wurden für die Modellkalibrierung, Vorhersage und anschließende Kartierung der Gesamtabundanz und des Artenreichtums der Regenwurmgemeinschaften sowie der räumlichen Verteilung der Arten verwendet.

2.1.4 Endgültige Variablenauswahl für die Modellkalibrierung (Arbeitsschritte 2 bis 4)

Die Modellkalibrierung umfasst die Auswahl geeigneter Prädiktoren, die unabänderlich die Genauigkeit der Modelle bestimmen, wobei Probleme der Variablen-Autokorrelation und -Multikollinearität die Modellleistung und -genauigkeit beeinflussen (Guisan & Zimmermann, 2000, Mod *et al.*, 2016; Salako, 2020). Die hier vorgenommene Modellkalibrierung umfasste drei Hauptschritte: 1) Identifizierung von Autokorrelation und Multikollinearität zwischen den Prädiktorvariablen, 2) Entfernung der stark korrelierten Variablen und 3) Auswahl der besten Kandidatenvariablen für die Modellierung.

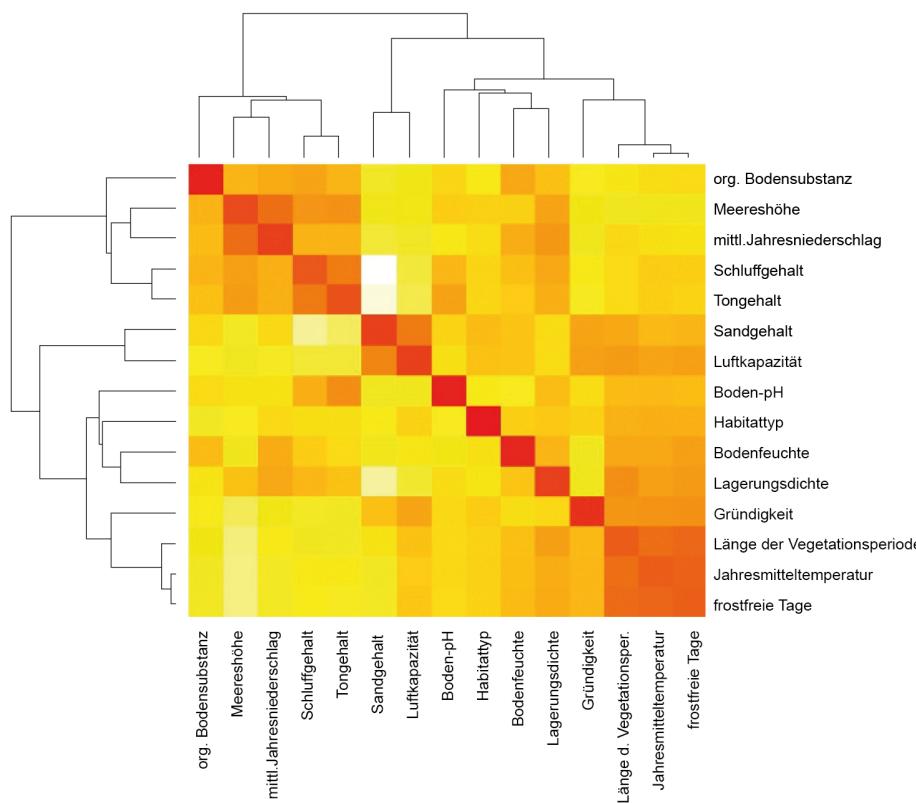
Um zunächst die numerischen Prädiktorvariablen auf Autokorrelation zu testen, wurden mit Hilfe einer Hauptkomponentenanalyse (PCA, mit der Funktion `pca` im R-Paket "FactoMineR"; Lê *et al.*, 2008) (visuell) autokorrelierte Variablen identifiziert, und die Pearson-Korrelationskoeffizienten zwischen Variablenpaaren mit der Funktion "ggpairs" im R-Paket Ggally berechnet. Zuvor wurden alle numerischen Variablen auf gemeinsame Mittelwerte (Z-Score von 0) und Einheitsvarianz standardisiert. Um signifikante Korrelationen zu identifizieren, wurde der Korrelationsschwellenwert auf $R^2 \geq 0,7$ gesetzt (Johnson *et al.*, 2002; Bobrowski *et al.*, 2021). Die Ergebnisse der PCA und der Korrelationsmatrix (Abbildung 3 bis Abbildung 5) identifizierten als autokorrelierte Variablen: die mittlere Jahrestemperatur ("Bio 1"), die stark mit der durchschnittlichen Anzahl der Tage pro Jahr über 0 °C und über 10 °C ("GDD0" bzw. "NGD10") korreliert ist, und den Gesamtjahresniederschlag ("Bio 12") mit der Meereshöhe. Deshalb wurden die Meereshöhe und die Anzahl der Tage pro Jahr über 0 °C und 10 °C (NGD10 und GDD0) aus den Modellen entfernt. Beibehalten wurden der mittlere Jahresniederschlag und die Jahresmitteltemperatur der Luft (Bio1 und Bio12) aufgrund ihrer physiologischen und ökologischen Bedeutung für Regenwürmer (Dormann *et al.*, 2013). Dadurch wurde die Zahl der Umweltvariablen auf 13 reduziert. Außerdem zeigte der Pearson-Korrelationskoeffizient, dass Sand- und Schluffgehalt des Bodens den gewählten Schwellenwert für die Autokorrelation von $R^2 \geq 0,7$ verletzten. Um die Multikollinearität zu testen, wurden anschließend während der Modellkalibrierung die Varianzinflationsfaktoren (VIF) berechnet mit einem Schwellenwert von $VIF > 5$, um Multikollinearität zwischen Prädiktoren zu erkennen (Dormann *et al.*, 2013). Auf der Grundlage der VIF-Ergebnisse wurde der Sandgehalt entfernt (Schluff- und Tongehalt wurden jedoch beibehalten), so dass 11 nicht korrelierte Umweltprädiktoren übrigblieben, die bei der endgültigen Anpassung der Modelle verwendet wurden (Tabelle 3).

Abbildung 3: Hauptkomponentenanalyse (PCA) der (standardisierten) quantitativen, d. h. numerischen vorausgewählten Prädiktorvariablen. Prozentwerte in den Achsenbeschriftungen geben den Anteil an Datenvariabilität an, die durch die Achse repräsentiert wird. „Cos²“ = Korrelationskoeffizienten im PCA-Raum



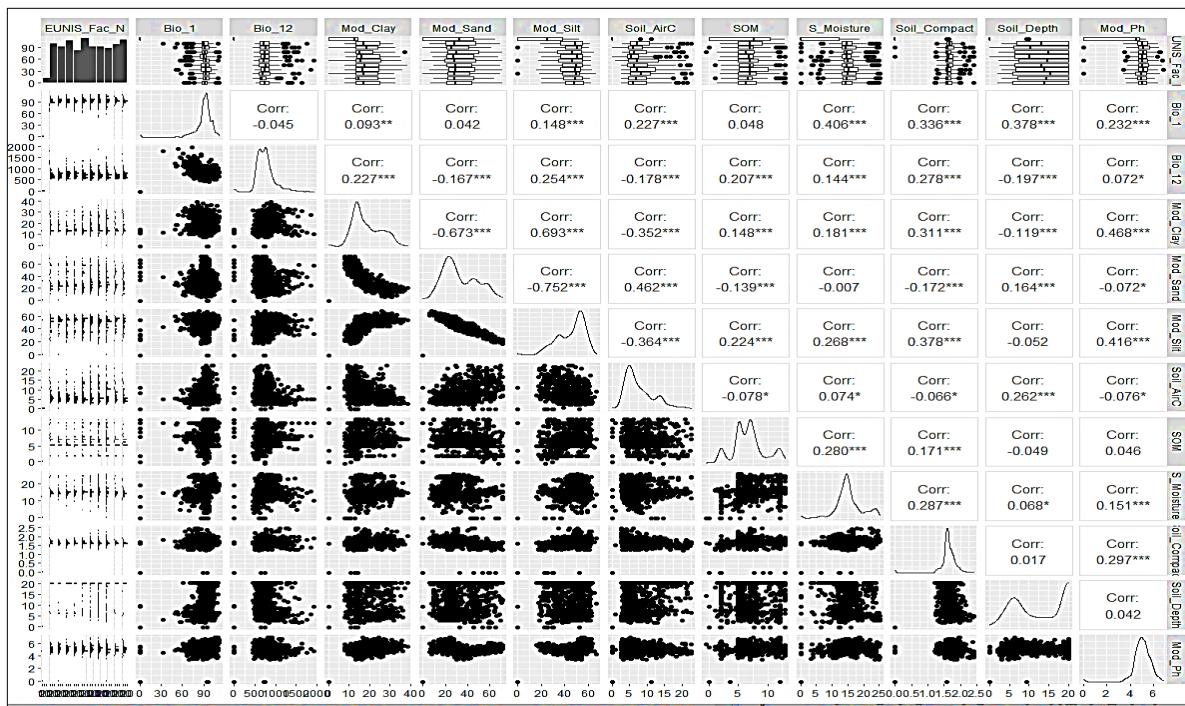
Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 4: Korrelations-Heatmap der 15 ursprünglich vorausgewählten standardisierten quantitativen Prädiktorvariablen. Eine dunklere Farbe (in Richtung Rot) zeigt eine höhere Korrelation zwischen Variablenpaaren.



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 5: Korrelationsmatrix der 13 standardisierten quantitativen Prädiktorvariablen nach Prädiktoreliminierung, basierend auf PCA- und Korrelationsanalyse



Quelle: Eigene Darstellung, Senckenberg Museum

Tabelle 4: Liste ausgewählter Umweltparameter, die nach Autokorrelations- und Multikollinearitätstests als Prädiktorvariablen in die Modellalgorithmen eingehen

Code	Beschreibung	Räumliche Auflösung	Einheit
Bio 1	Jahresmitteltemperatur	1 km	°C (Beachte: im Datensatz °C*10)
Bio 12	Mittlerer Jahresniederschlag	1 km	mm
Mod_clay	Tongehalt des Bodens	250 m	Prozent (%)
Mod_silt	Schluffgehalt des Bodens	250 m	Prozent (%)
SoilAirC	Luftkapazität des Bodens	250 m	Volumen- Prozent (Vol.-%)
SOM	Gehalt an organischer Substanz des Bodens	250 m	Prozent (%)
ResMsk_SM/S_Moisture	Bodenfeuchte	250 m	% Bodenwasser- gehalt in 2 – 5 cm Tiefe
Soil_compact	Lagerungsdichte des Bodens	250 m	Trockenrohdichte g cm ⁻³
Soil_depth	Gründigkeit des Bodens	250 m	Tiefe in cm
Mod_pH	pH (0,01 M CaCl ₂)	250 m	dimensionslos
EUNIS_Hab	Habitat-/Landnutzungstyp	100 m	nominalskaliert (s. Tabelle 3)

2.1.5 Modellkalibrierung und -anpassung (Arbeitsschritte 5 und 6)

2.1.5.1 Beschreibung der Modellalgorithmen

Alle Analysen wurden in der Programmiersprache R (R Core Team, 2020) durchgeführt. Vier Algorithmen wurden ausgewählt, um die Modelle anzupassen (Kalibrierung zwischen bekannten Regenwurmvorkommen und Prädiktorvariablen) und anschließend die endgültigen Verteilungsvorhersagen des Modells zu erstellen. Bei den Algorithmen handelte es sich einerseits um traditionelle Regressionsmodelle (verallgemeinerte lineare Regressionsmodelle [GLM], verallgemeinerte additive Modelle [GAM]) und andererseits um Algorithmen des maschinellen Lernens (generalized boosted regression-tree models [GBM] und Random Forest-Modelle [RF]). Diese Modellalgorithmen sind in der Literatur mehrfach ausführlich beschrieben worden (z. B. Guisan und Zimmermann, 2000; Li & Wang 2013; Guisan *et al.*, 2017); Tabelle 4 enthält eine Zusammenfassung ihrer grundlegenden Merkmale und Eigenschaften.

2.1.5.2 Modellanpassungen für GLM, GAM, GBM und RF

Generalized linear models (GLM) wurden in der Basisversion von R unter Verwendung einer Verknüpfungsfunktion der Gauß-Familie erstellt, *Generalized additive models* (GAM) in R mit dem Paket "themgcv". GLMs können sowohl für die Vorhersage von binären Daten (mit den Werten 1 und 0, d. h. Vorhandensein oder Fehlen) als auch von numerischen Daten (1, 2, 3, ...; d. h. Gesamtabundanzen oder Artenreichtum) verwendet werden (Fox, 2015; Bhatt *et al.*, 2017; Capinha *et al.*, 2018; Li & Wang, 2013). Sowohl für die Anpassung des GLM als auch des GAM wurde zunächst das Histogramm der ausgewählten Prädiktorvariablen untersucht, um ihre Verteilung zu überprüfen, da für die lineare Regressionsmodellierung normalverteilte Werte erforderlich sind (Mangiafico, 2016, Liu *et al.*, 2018). Der mittlere Jahresniederschlag ("Bio 12") und die Luftkapazität des Bodens wiesen schiefe Verteilungen auf und wurden daher log10-transformiert. GLM-Modelle wurden mit der Gesamtabundanz der Gemeinschaft und dem Artenreichtum sowie den Daten über die An- und Abwesenheit einzelner Arten (P/A; für die Artenverteilung) als Responsevariablen und mit den 11 oben genannten Umweltvariablen als Prädiktoren (erklärende Variablen) angepasst. Die Modelle wurden unter Verwendung der folgenden R-Funktionsaufrufe formuliert, wobei Polynome zweiter Ordnung ("poly"; zur Berücksichtigung nichtlinearer Beziehungen zwischen Regenwürmern und Prädiktoren) und eine Glättungsfunktion ("s") zur Gleichung für polynomiale Regression in GLM und Glättung in GAM hinzugefügt wurden:

Lineare Regression: $Y = \text{glm}(a \sim x_0 + x_1 + x_2 + x_3 + x_4 + \dots + x_n)$,

Polynomiale Regression: $Y_1 = \text{glm}(a \sim \text{poly}(x_0, 2) + \text{poly}(x_1, 2) + \text{poly}(x_2, 2) + \text{poly}(x_3, 2) + \text{poly}(x_4, 2) + \dots + \text{poly}(x_n, 2))$ [Beachte: polynomiale Funktionsaufrufe schließen den linearen Term ein]

Additive Modelle: $Y_2 = \text{gam}(a \sim s(x) + s(x_1) + s(x_2) + s(x_3) + \dots + s(x_n))$

wobei Y, Y1, Y2 die vorhergesagten Responsevariablen (Gesamthäufigkeit, Artenreichtum, Wahrscheinlichkeit des Auftretens einzelner Arten), a die beobachtete Responsevariable, die bei der Kalibrierung des Modells verwendet wurde, x die Umwelt-/Prädiktorenvariablen und s die Glättungsfunktion darstellen. „glm“ ist die R-Funktion für ein *generalized linear model* und gam die Funktion für ein *generalized additive model*. Wir wählten die besten Modelle in einer schrittweisen (Prädiktor-)Variablenauswahl aus, die auf der Grundlage des Akaike-Informationskriteriums (AIC) jedes getesteten Modells (= Prädiktorenkombination) durchgeführt wurde. Es wurden die Modelle mit dem niedrigsten AIC ausgewählt.

Tabelle 5: Zusammenfassung der vier getesteten Algorithmen zur Modellierung der räumlichen Verbreitung von Regenwürmern in Deutschland

Modell-klasse	Me-thode	Beschreibung	Besondere Eigen-schaften und Funktio-nen
Klassische Regression	GLM	Generalisierte lineare Regression (GLM) ist eine Verallgemeinerung allgemeiner linearer Modelle (McCullagh & Nelder, 1989). Sie ermöglichen die Anpassung eines linearen Modells zwischen erklärenden (Prädiktor-)Variablen und der Responsevariablen über eine Verknüpfungsfunktion (basierend auf dem Datenverteilungstyp) und liefert die Größe der Varianzen zwischen den Responsevariablen und die statistische Anpassung an die Prädiktorvariablen.	Weniger komplex, häufig verwendet und kann mit allen Arten von Daten (binär, kontinuierlich, kategorial, Zählung) angepasst werden. Gut geeignet zur Modellextrapolation.
Klassische Regression (Glättung)	GAM	<p>Verallgemeinerte additive Modelle (GAMs) sind nicht-parametrische Erweiterungen von GLMs und bieten somit das Potenzial für bessere Anpassungen an heterogene Daten als GLMs (Hastie & Tibshirani, 1986). GAMs verwenden datendefinierte Glättungsfunktionen, um nichtlineare Art-Umwelt-Beziehungen anzupassen. Die Glättungsfunktionen werden für jede erklärende Variable unabhängig berechnet und dann hinzugefügt, um das endgültige Modell zu erstellen.</p> <p>Wird zum Anpassen nichtlinearer Funktionen verwendet. Baut auf dem verallgemeinerten linearen Modell auf, indem der lineare Prädiktor durch einen additiven Prädiktor ersetzt wird (Hastie & Tibshirani, 1986), der die Modellierung komplexerer ökologischer Reaktionen erlaubt (z. B. Elith <i>et al.</i>, 2006).</p>	<p>Kann in manchen Fällen besser abschneiden als GLM, wenn komplexe Arten-Umweltbeziehungen betrachtet werden.</p> <p>Ebenso gut zur Modellextrapolation geeignet.</p>
Maschinelles Lernen (Boosting und Regression)	GBM	Generalized Boosting Models (GBM), auch unter der Bezeichnung Boosted Regression Trees (BRT) bekannt, verwenden einen Boosting-Algorithmus, der den Regression-Tree-Algorithmus iterativ verwendet, um eine Kombination von Bäumen in einem schrittweisen Vorwärtsverfahren zu konstruieren. Es konzentriert sich auf die schwächsten Teile des bisher erstellten Modells, indem jeder neue Baum an das Residuum der zuvor angepassten Bäume angepasst wird (Elith <i>et al.</i> , 2006).	Gut zum Lösen von Klassifizierungs- und Regressionsproblemen mit kontinuierlichen und kategorialen Daten (Elith <i>et al.</i> , 2006; Li & Wang, 2013)
Maschinelles Lernen (Klassifikation und Regression with Bagging)	RF	<p>Random Forest (RF) ist eine Ensemble-Methode, die viele verschiedene Klassifikations-(Entscheidungs-)Bäume beinhaltet und Breimans Random-Forest-Algorithmus zur Auswahl der besten Regression und Klassifikation implementiert (Breiman, 2001).</p> <p>Bagging ist ein charakteristisches Merkmal von RF (wodurch es sich von GBM unterscheidet): Es ist ein Ensemble-Algorithmus, der mehrere Modelle mit verschiedenen Teilmengen eines Trainingsdatensatzes anpasst und dann die Vorhersagen aller Modelle kombiniert. Das Prinzip hinter dem Bagging in RF besteht darin, mehrere Bäume an verschiedene Resampling-Teilmengen der ursprünglichen Datensätze anzupassen und dann die Ergebnisse der verschiedenen Teilstichproben zu mitteln.</p>	Hohe Vorhersageleistungen (Guisan <i>et al.</i> , 2017)

Die Modellanpassung (Kalibrierung) für GBM wurde mit den Paketen DISMO (Version 1.3-5, Hijmans *et al.*, 2019) und BIOMOD 2 (Version 3.5.1, Thuiller, *et al.*, 2008, 2021) in R durchgeführt, um alle Einzelfunktionen, die mit jedem der Pakete verbunden sind, optimal zu nutzen. Bestimmte Parameter sind bei der GBM-Modellanpassung wichtig, z. B. die Gesamtzahl der anzupassenden Bäume, die Lernrate oder Schrumpfung, die Anzahl der Kreuzvalidierungen und die maximale Tiefe der Variableninteraktion (Elith *et al.* 2008; Guisan *et al.*, 2017). Nach einer Reihe von Modellläufen, bei denen diese Parameter optimiert wurden, wurde das endgültige Modell für die Vorhersage der Regenwurm-Gemeinschaft/Artabundanz mit den 11 Prädiktoren mit 1000 Bäumen (n.trees) angepasst, die nacheinander hinzugefügt wurden, um eine Überanpassung zu vermeiden. Die Lernrate (lr) wurde auf 0,001 gesetzt und ein bag.fraction (siehe Erläuterung in Tabelle 5) von 0,5 mit einer maximalen Kreuzvalidierung von 10 Iterationszyklen (cv.folds =10) verwendet. Die endgültige Auswahl basierte auf dem Modell mit der geringsten Vorhersageabweichung. Die folgende Funktion wurde für die GBMs verwendet:

```
Y = gbm.step(data=DF, gbm.x =, gbm.y =, family = "gaussian" or "Poisson", tree.complexity = 2, learning.rate = 0.001, bag.fraction = 0.5, cv.folds=10)
```

wobei DF die Trainings-/Testdaten (Datenrahmen, der sowohl die Antwort- als auch die Prädiktorvariablen kombiniert), gbm.x die Prädiktorspalte(n), gbm.y die Responsespalte und tree.complexity die Anzahl der Bäume darstellt.

Das Random-Forest-Modell wurde mit Hilfe des R-Pakets RandomForest (Liaw & Wiener, 2002) mit den 11 Umweltvariablen als Prädiktoren und den Reaktionsvariablen der Gemeinschaftsgesamtabundanz und des Artenreichtums sowie den artspezifischen Beobachtungsdaten (P/A) angepasst. Die Mindestanzahl der Bäume wurde auf 1000 und die Knotengröße auf 10 festgelegt, was die optimale Einstellung darstellt, die die besten Anpassungsstatistiken ergibt. Das RF-Modell wurde mit der folgenden Funktion formuliert:

```
Y = randomForest(x= n[,5:15], y= n[,3], ntree = 1000,nodesize = 10, importance = T)
```

wobei x die Spalten des Datenrahmens mit den Prädiktoren, y die Spalte des Datenrahmens mit der Responsevariablen und ntree die Anzahl der Bäume darstellt.

Die Anpassung des RF-Modells und die Vorhersagen wurden auf zwei Arten durchgeführt: (1) "Regression", bei der die Responsevariablen der Gesamtabundanz und des Artenreichtums oder die Vorkommensdatensätze von Anwesenheit (1) und Abwesenheit (0) für die Vorhersage als numerisch behandelt wurden, und (2) "Klassifizierung", bei der die Responsevariablen als Faktoren oder kategoriale Variablen behandelt wurden, um eine binäre Vorhersage von Anwesenheit (1) oder Abwesenheit (0) zu erstellen (Hijmans & Elith, 2019; Valavi *et al.*, 2021). Es wurden eine Modellplausibilitätsprüfung sowie eine Variablenbeitrags- oder -bedeutungsanalyse sowohl für GBM als auch für RF durchgeführt, mithilfe einer Reaktionskurve (partieller Abhängigkeitsplot), die die Reaktionen der Regenwürmer auf die ausgewählten Umweltvariablen zeigt.

Zur Modellierung und Kartierung des Regenwurm-Artenreichtums können sowohl makroökologische Modelle (*macro-ecological models*, MEM; Regressionsmodelle mit bekanntem Artenreichtum als Responsevariable) als auch gestapelte Artenverteilungsmodelle (*stacked species distribution models*, SSDM; Kombination der Vorkommensvorhersagen einzelner Arten) verwendet werden. Die Entscheidung fiel auf ein MEM aufgrund seiner relativen Vorteile gegenüber SSDMs (Biber *et al.*, 2020): in erster Linie aufgrund der Tatsache, dass SSDMs den Gesamtartenreichtum nicht angemessen darstellen, wenn nicht alle Arten modelliert werden können (d. h. aufgrund unzureichender Daten für seltene Arten, wie im vorliegenden Fall). Beide Methoden wurden getestet, und es zeigte sich, dass SSDM den Artenreichtum im Vergleich zu MEM stark unterschätzt.

2.1.6 Prognosen (GIS-Kartenerstellung aus den Modellergebnissen) (Arbeitsschritt 7)

2.1.6.1 Visualisierung der Modellergebnisse

Alle Modellvorhersageraster wurden in die GIS-Umgebung importiert. Um später Modellierungsresultate genauer auf die Karten projizieren zu können, wurden die Rasterdateien anschließend in das für Europa passende Projektionssystem "ETRS89 LAEA" (EPSG-Code: 3035) umgewandelt und eine die Kartendarstellung verbessерnde Symbolisierung vorgenommen.

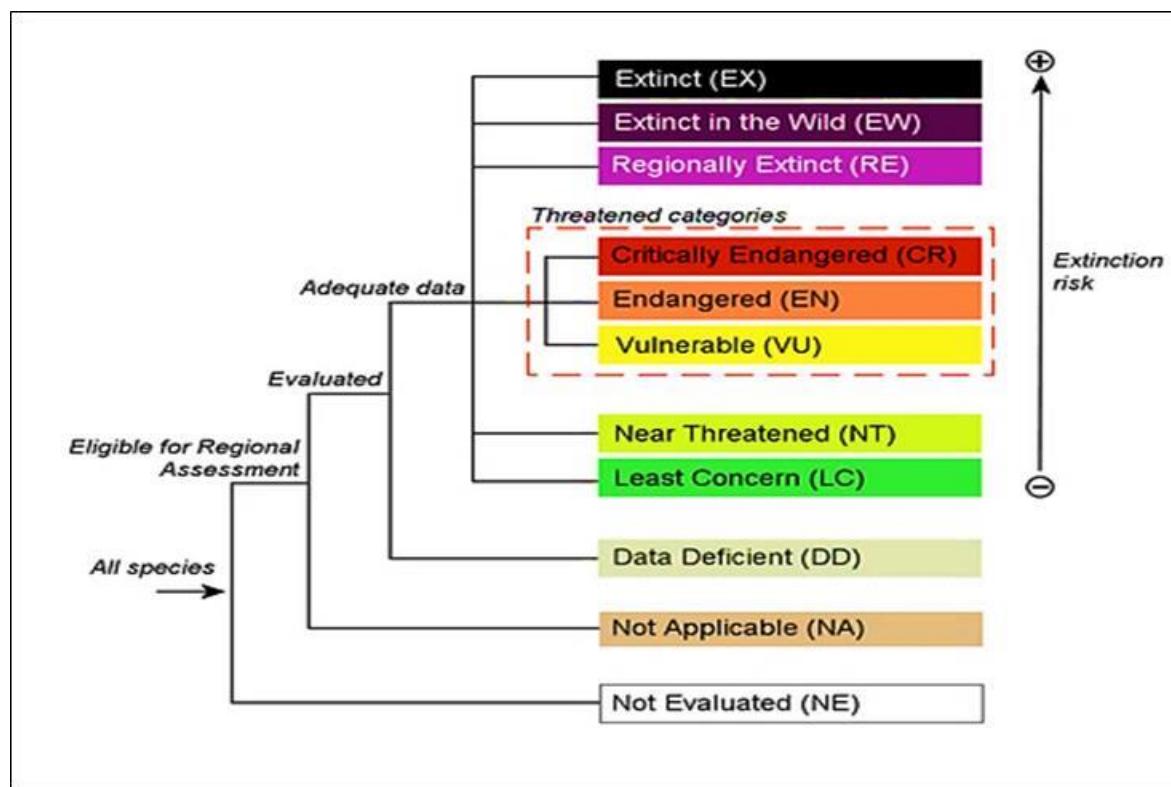
2.1.6.2 Diversitätsabschätzung

Die Gesamtabundanz einer Gemeinschaft und der Artenreichtum sind zwei Maße, die eine modellierte Regenwurmgemeinschaft beschreiben. Keiner der beiden Indikatoren beschreibt jedoch die Vielfalt einer beobachteten oder (modellierten) potenziellen Gemeinschaft adäquat. Üblicherweise verwendete nichtparametrische Diversitätsindizes wie der Shannon-Wiener- oder Simpson-Index oder die verschiedenen Evenness-Indizes (Magurran 2004) erfordern Abundanzdaten für alle konstituierenden Arten einer (lokalen) Gemeinschaft, die nicht für alle bei der Modellkalibrierung verwendeten Standorte verfügbar waren. Daher konnte die potenzielle räumliche Verteilung solcher Diversitätsindizes nicht angemessen modelliert werden. Zur Überwindung dieser Schwierigkeit und als erste Annäherung an die potenzielle Vielfalt der Regenwürmer wurde eine GIS-Überlagerung der Ergebnisse der Modellierung der Gesamtabundanz und des Artenreichtums der Regenwurmgemeinschaft erstellt. Dazu wurden die GIS-Rasterdateien für die Gesamtabundanz und den Artenreichtum auf eine Skala von 1 bis 5 standardisiert (um die beiden Messgrößen vergleichbar zu machen) und der Skalenwert jeder entsprechenden Zelle in den Dateien multipliziert. Die sich daraus ergebenden Werte wiederum wurden in eine fünfstufige Skala der potenziellen Vielfalt ("sehr gering", "gering", "mittel", "hoch" und "sehr hoch") kategorisiert, die dann mit denselben GIS-Verfahren wie für die Gesamtabundanz und den Artenreichtum kartiert wurde. Da dies nur eine sehr grobe Kombination aus Gesamtabundanz und Artenreichtum darstellt, können die Ergebnisse nur als eine erste Annäherung an die potenzielle Lebensraumeignung für die Regenwurmvielfalt betrachtet werden.

2.1.6.3 Gefährdungsabschätzung

Um die Größe des geografischen Verbreitungsgebiets der Arten für eine schnelle und rasche Bewertung ihrer Erhaltungsbedürfnisse und ihres Status auf nationaler Ebene zu bestimmen, wurden die Größe des Verbreitungsgebiets (AOO) der Arten mithilfe des Rangemap-Pakets in R (Version 0.1.19, 2021; Cobos *et al.*, 2021) geschätzt. Hierfür wurde ein Datenrahmen aus den (bereinigten) Vorkommensaufzeichnungen der 12 ausgewählten Arten in einem 2 x 2 km-Raster erstellt, wie von der IUCN für den räumlichen Maßstab empfohlen, der für die Bestimmung von AOO unter Aufrechterhaltung der Konsistenz erforderlich ist (IUCN 2022). Um EOO zu kartieren, wurde die Funktion „Range Map Plot“ des Rangemap-Pakets verwendet, mit der Pufferentfernung von 5000 m und einem konvexen Hüllpolygon. Die Ergebnisse der Arten wurden anhand der in km² gemessenen EOO- und AOO-Werte in die IUCN-Gefährdungsklassen eingeteilt (Abbildung 6). Die Schwellenwerte der IUCN für EOO sind 100 km², 5000 km² und 20 000 km² für vom Aussterben bedrohte, sehr gefährdete und gefährdete Arten, während die entsprechenden Schwellenwerte für AOO 10 km², 500 km² und 2000 km² betragen (Gaston & Fuller, 2009; IUCN, 2022). Sowohl die vorhergesagten Verbreitungskarten als auch die geografischen Verbreitungsgrößen erlaubten, Arten räumlichen Verbreitungsgebieten zuzuordnen: (1) großes Verbreitungsgebiet – Arten mit weiter Verbreitung und AOO > 2000 km², (2) mittleres Verbreitungsgebiet – Arten mit AOO > 1000 km², (3) eingeschränktes oder kleines Verbreitungsgebiet – Arten mit AOO < 500 km² und endemische Arten oder Arten mit besonderem Habitat mit AOO < 200 km².

Abbildung 6: Struktur und Entscheidungsfluss der IUCN-Kategorien zur Gefährdungsbeurteilung von Arten



Quelle: IUCN, 2022

2.1.7 Model assessment/evaluation (Arbeitsschritt 8)

Zur Bewertung der Modellleistung wurden verschiedene Bewertungsmetriken und -methoden verwendet, um Fallstricke und mögliche Verzerrungen zu vermeiden, die mit der Verwendung einer einzelnen Methode wie der weit verbreiteten „area under the receiver operating characteristic curve“ (AUC_{ROC}) verbunden sein können. (Jiménez-Valverde 2011).

Zunächst wurde die Methode der geteilten Stichprobe oder Kreuzvalidierung (CV) verwendet, bei der die Vorhersageleistung des Modells durch Aufteilung der Daten der Responsevariablen in einen Trainings- und einen Testsatz im Verhältnis 70:30 bewertet wird (Phillips *et al.*, 2008; Hijmans & Elith 2019; Guisan *et al.*, 2017). Diese Methode, bei der die Stichprobendaten nach dem Zufallsprinzip ausgewählt werden, hat sich auch als geeignet erwiesen, die Auswirkungen der räumlichen Autokorrelation in ökologischen Datensätzen zu verringern (Roberts *et al.*, 2017; Valavi *et al.*, 2019; Neftalí & Márcia, 2020). Die Trainingsdatensätze werden zur Anpassung (Kalibrierung) des Modells verwendet, während die Testdaten zur Bewertung der Vorhersageleistung des Modells genutzt werden.

Quantitative (numerische) Daten, wie z. B. die Gesamtabundanz einer Gemeinschaft oder die Werte für den Artenreichtum, lassen sich am besten durch konventionelle Statistiken wie R^2 , den mittleren quadratischen Fehler, u. ä. bewerten (Guisan & Zimmerman, 2000; Guisan *et al.*, 2017). Die Beziehung zwischen der beobachteten Gesamtabundanz bzw. dem Artenreichtum und den für sie vorhergesagten Werten wurde mithilfe des Bestimmtheitsmaßes (R^2) und dem Konkordanzindex (Anteil der übereinstimmenden Paare geteilt durch die Gesamtzahl der möglichen Bewertungspaare; ein Äquivalent zu AUC_{ROC} ; Harell *et al.*, 1982; Schmid *et al.*, 2016) untersucht. Außerdem wurden Streudiagramme erstellt, um die Beziehungen zwischen den beobachteten

und vorhergesagten Werten zu visualisieren, und eine lineare Regression zwischen den beiden Datensätzen berechnet.

Für die Modellierung der Verteilung der Vorkommenswahrscheinlichkeiten von Arten unter Verwendung von binären Präsenz-/Absenzdaten (PA) bewerteten wir die Vorhersagefähigkeit der Modelle mithilfe der schwellenunabhängigen Statistik AUC_{ROC} , die im R-Paket pROC (Version 1.18, 2021, Robin *et al.*, 2011) implementiert ist. AUC_{ROC} wird häufig zur Bewertung der Modellleistung in der Ökologie verwendet; sie bewertet die Fähigkeit eines Modells, zwischen Standorten mit Vorhandensein oder Nichtvorhandensein einer Art zu unterscheiden (Guisan *et al.*, 2017; Valavi *et al.*, 2021) und reicht von 0 bis 1, wobei eine AUC von 0,5 (oder niedriger) eine nicht besser als zufällige Vorhersage beschreibt, 0,7 bis 0,8 als akzeptabel, 0,8 bis 0,9 als ausgezeichnet und mehr als 0,9 als hervorragend gewertet wird (Manel *et al.*, 2001; Guisan *et al.*, 2017).

Außerdem wurden mit dem R-Paket "caret" (Version 6.0-91, Kuhn 2008, 2022) schwellenabhängige (Konfusionsmatrix-)Indizes berechnet, um die Vorhersagegenauigkeit der Modelle zu bestimmen. Cohens Kappa-Koeffizient und andere Maße (Sensitivität, Spezifität, Gleichgewichtsgenauigkeit) sowie die Kappa-Statistik wurden verwendet, um die Diskriminierung und die Leistungen der Modelle auf der Grundlage der Präsenz-/Absenzdaten zu bewerten. Cohens Kappa wurde mit der folgenden Formel berechnet:

$$K = (Po - Pe) / (1 - Pe)$$

Dabei steht K für den Kappa-Koeffizienten, Po für die Gesamtgenauigkeit des Modells und Pe für die Übereinstimmung zwischen den Modellvorhersagen und den tatsächlichen Klassenwerten, als ob diese zufällig auftreten. Die Kappa-Werte reichen von -1 bis 1, mit 0 (keine Übereinstimmung) und 1 (perfekte Übereinstimmung); in wenigen Fällen können die Kappa-Werte einen negativen Wert (< 0) annehmen, was bedeutet, dass die Übereinstimmung geringer ist als zufällig zu erwarten wäre.

Darüber hinaus wurde die Plausibilität der Modelle und ihre ökologische Relevanz überprüft, indem der quantitative Beitrag jeder Prädiktorvariablen zu den Modellergebnissen ermittelt wurde sowie durch visuelle Inspektion der Responsekurven der Responsevariablen (Gesamtabundanz oder Artenreichtum der Regenwürmer, Vorkommenswahrscheinlichkeit der Arten) in Bezug auf jede numerische Umweltvariable und (Allouche *et al.*, 2006; Guisan *et al.*, 2017). Responsekurvendiagramme beschreiben das mechanistische Nischenverhalten einer Art (in Bezug auf das Vorkommen oder die Populationsgröße) gegenüber ausgewählten Umweltgradienten und können die Gradientenbereiche aufzeigen, in denen eine Art ihre optimale Fitness ("Präferenz") und/oder Schwellenwerte erreicht, bei denen sich das Reaktionsverhalten der Art ändert. Dies ist wichtig, um zu beurteilen, wie ökologisch plausibel eine Vorhersage auf der Grundlage des vorhandenen Wissens über das Verhalten der Art sein ist (Hijmans & Elith, 2019). Random Forest erzeugt insbesondere für jede Prädiktorvariable eine partielle Abhängigkeitsresponsekurve, die alle anderen Prädiktoren berücksichtigt.

Da es sich bei den hier verwendeten Daten zur Gesamtabundanz und zum Artenreichtum um quantitative Responsevariablen handelt, wurde auch das Bestimmtheitsmaß der Regression (R^2) zwischen den Beobachtungen (Funddaten) zur Abundanz und zum Artenreichtum und ihren vorhergesagten Werten bestimmt und die lineare Regressionsgleichung berechnet. Dieses Maß für die Anpassungsgüte ist einer der am häufigsten verwendeten Ansätze zur Bewertung der Vorhersagegüte eines Modells, insbesondere bei quantitativen ökologischen Daten. Außerdem wurde der mittlere quadratische Fehler (RMSE) anhand der Korrelationskoeffizienten berechnet (Guisan & Zimmerman, 2000; Guisan *et al.*, 2017). Zur visuellen Überprüfung wurden Streudiagramme erstellt, in denen die vorhergesagten Werte und die ursprünglichen Beobachtungsdaten mit der ggplot-Funktion aus dem R-Paket "ggplot2" in (Wickham, 2016) verglichen wurden.

3 Ergebnisse

3.1 Modellperformanz und endgültige Modellauswahl

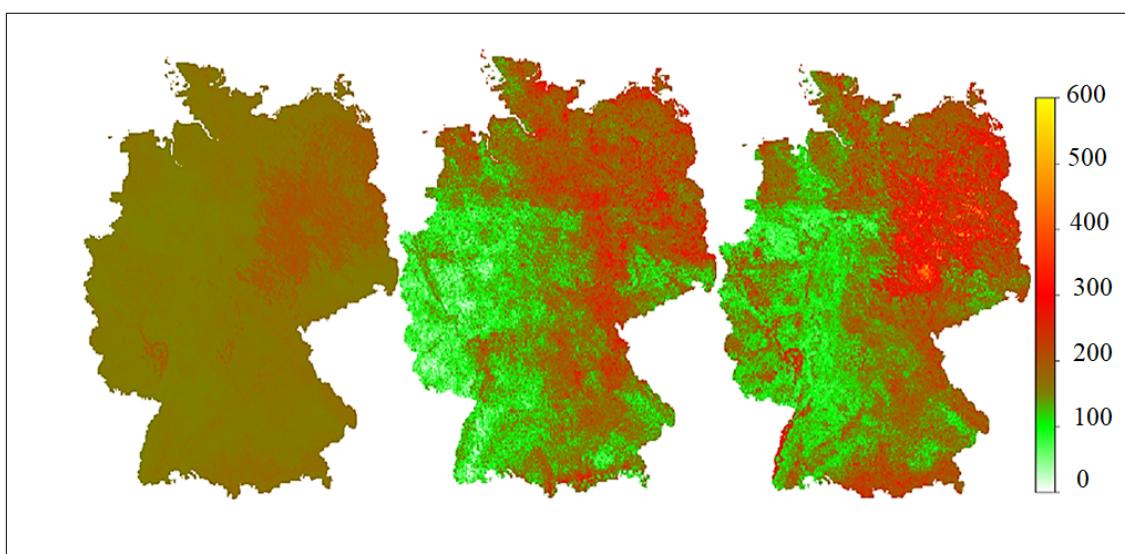
Die Bewertung der Leistung der getesteten Modelle zur Vorhersage des Artenreichtums und der Gesamtabundanz der Regenwurmgesellschaften zeigt, dass GLM die geringste Vorhersagegüte erreichen (Tabelle 6), sie zeigen sowohl die niedrigsten R^2 - als auch C-Indexwerte. Obwohl die räumliche Vorhersage des Artenreichtums und der Gesamtabundanz von GBM und RF sehr ähnlich aussehen (Abbildung 7, Abbildung 8), unterscheiden sich ihre R^2 - und C-Indexwerte sehr. RF ergab den höchsten R^2 -Wert für die Gesamtabundanz (0,840) und Artenreichtum (0,574), während die R^2 -Werte des GBM weit darunter liegen (0,155 bzw. 0,115). Die C-Indexwerte sind ebenfalls höher für RF als für GBM sowohl für Gesamtabundanz (0,861 gegenüber 0,657) als auch Artenreichtum (0,819 gegenüber 0,658).

Tabelle 6: Werte für Bestimmtheitsmaß und Übereinstimmungskoeffizient (C-Index) der unterschiedlichen Modellalgorithmen zur Vorhersage der räumlichen Verbreitung des Artenreichtums und der Gesamtabundanz der Regenwurmgesellschaften

Biodiversitätsmaß	Modell	R^2	C-Index
Gesamtabundanz	GLM	0,089	0,603
	GBM	0,155	0,657
	RF	0,840	0,861
Artenreichtum	GLM	0,029	0,583
	GBM	0,115	0,658
	RF	0,574	0,819

Abbildung 7: Initiale Gesamtabundanzvorhersage für Regenwurmgesellschaften durch drei unterschiedliche Modellalgorithmen

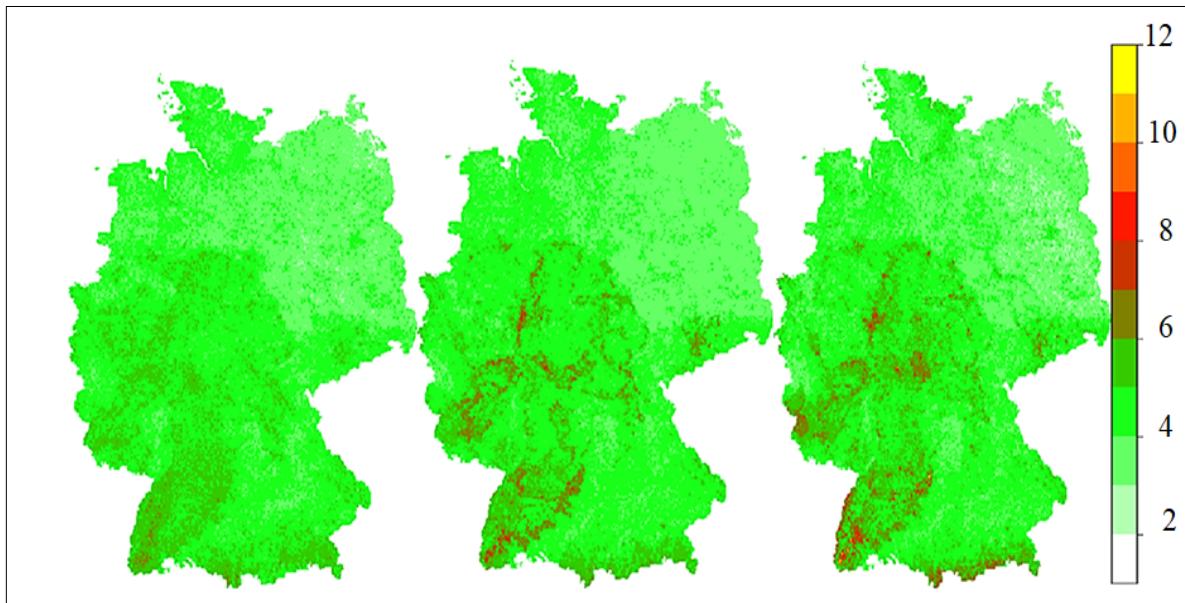
links: GLM – Mitte: GBM – rechts: RF



Quelle: Eigene Darstellung, Senckenberg Museum

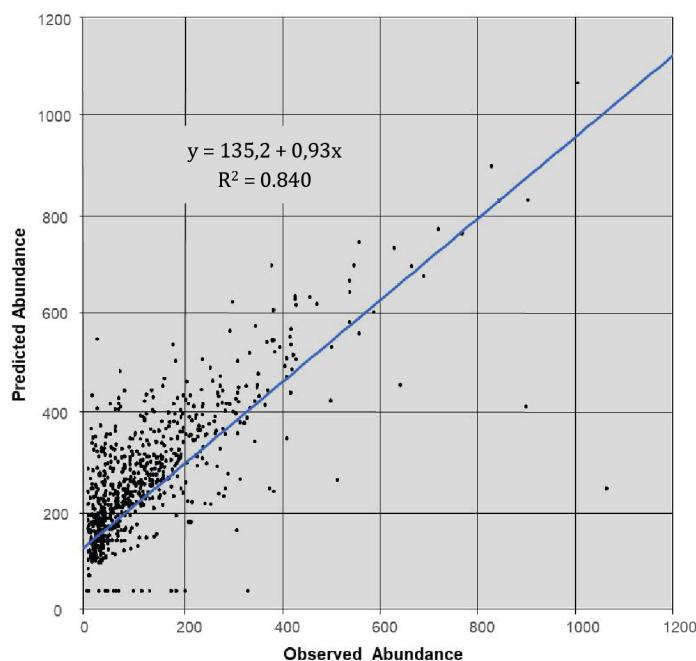
Abbildung 8: Initiale Artenzahlvorhersage für Regenwurmgesellschaften durch drei unterschiedliche Modellalgorithmen

links: GLM – Mitte: GBM – rechts: RF



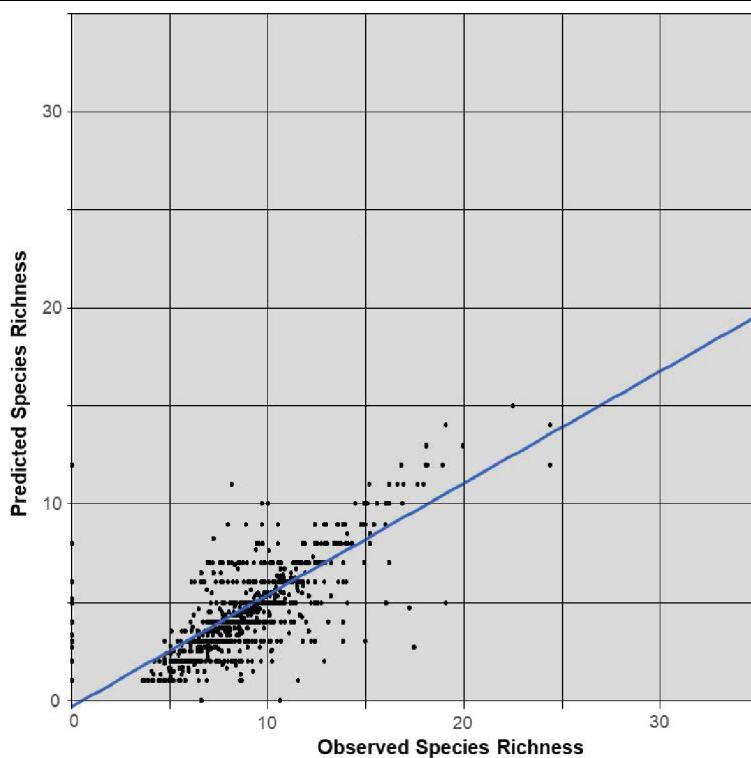
Quelle: Eigene Darstellung, Senckenberg Museum

Schließlich zeigen auch Streudiagramme und Regressionen der vorhergesagten und der beobachteten (d. h. originalen) Werte die hohe Vorhersagekraft der Random Forest-Modelle. Während die Übereinstimmung der modellierten mit den bekannten Werten der Gesamtabundanz der Regenwürmer (Abbildung 9) sehr gut ist, unterschätzen die Modelle den Artenreichtum der Regenwurmgesellschaften leicht (Abbildung 10).

Abbildung 9: Streudiagramm der vorhergesagten vs. beobachteten Gesamtabundanz von Regenwurmgesellschaften (finales Random Forest-Modell)

Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 10: Streudiagramm der vorhergesagten vs. beobachteten Artenzahl von Regenwurmgesellschaften (finales Random Forest-Modell)



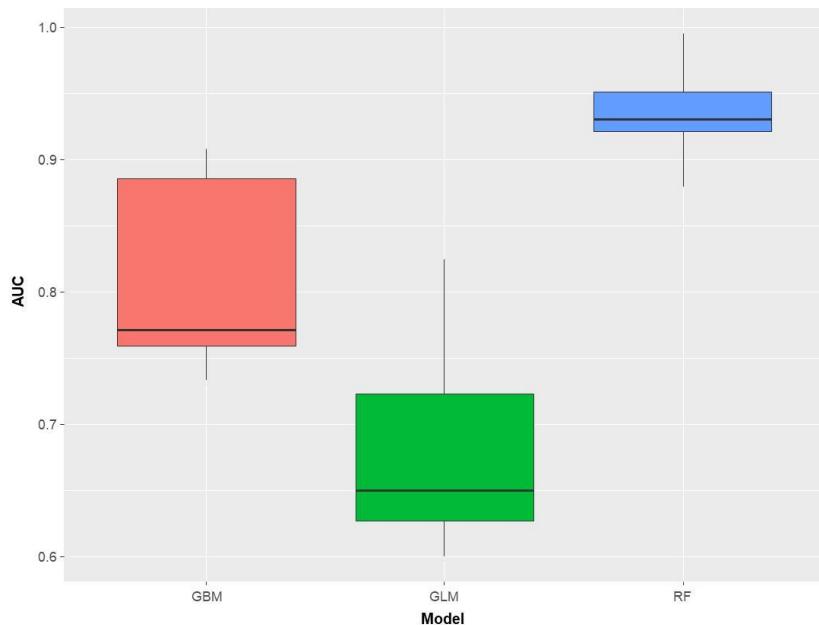
Quelle: Eigene Darstellung, Senckenberg Museum

Für die Vorhersagen der Vorkommenswahrscheinlichkeiten der ausgewählten Arten wurden die AUC_{ROC} - und Cohen-Kappa-Werte aller Modelle bestimmt und die Mittelwerte für alle Arten berechnet. Auf die GAMs wurde im Weiteren verzichtet, da sie keinen signifikanten Unterschied in ihrer Leistung im Vergleich zu den GLMs erkennen ließen. Die Ergebnisse der Modellvorhersage für die drei verbleibenden Algorithmen zeigen mittleren AUC_{ROC} - und Cohen-Kappa-Werte im Bereich von 0,65 – 0,97 bzw. 0,12 – 0,90 (Abbildung 11 und Abbildung 12). Beim Vergleich der kombinierten AUC_{ROC} - und Cohen-Kappa-Werte (Abbildung 13) führt RF zu den höchsten mittleren AUC_{ROC} - und Kappa-Werten mit geringerer Variation im Vergleich zu den anderen Modellen.

Die Modellvorhersagen sind für alle Arten sehr gut (AUC -Bereich aller Modelle: 0,601 – 0,995; Kappa-Bereich: 0,096 – 0,982; Tabelle 7), was auf eine allgemein gute Übereinstimmung zwischen Modellvorhersage und wahren Beobachtungsdaten hinweist (Manel, Ceri Williams, & Ormerod, 2001; Salako, 2015). GLM hat die niedrigsten AUC -Werte mit einem Durchschnitt von 0,601, wobei drei Arten (*A. eiseni*, *A. limicola*, *L. badensis*) höhere GLM- AUC -Werte von 0,825, 0,775 bzw. 0,796 aufweisen. Die an Cohens Kappa bemessene GLM-Leistung ist mit einem mittleren Wert von 0,210 sehr schlecht. Dies spiegelt sich besonders in der Modellleistung für *A. longa* wider, die den schlechtesten Mittelwert (0,096) aller Arten zeigt. GBM schneidet zwischen allen Modellen mit einer mittleren Arten- AUC und einem Kappa von 0,807 bzw. 0,463 ab. Auch hier wurde der niedrigste GBM-Kappa-Wert für *A. longa* gefunden (0,146). RF zeigt bei allen Arten die höchsten mittleren AUC - und Kappa-Werte mit Mittelwerten von 0,940 bzw. 0,836. Wiederum wurde der niedrigste RF-Kappa-Wert für *A. longa* (0,357) ermittelt. Während alle Modellalgorithmen die besten Kappa-Werte für epigäische Arten zeigen (mittleres Kappa für GLM, GBM, RF: 0,309, 0,560 bzw. 0,891), erzielte RF die beste mittlere AUC für endogäische Arten (Mittelwert: 0,959), während GMB zur besten mittleren AUC für epigäische Arten (0,826) führte. Die GLM- AUC s unterscheiden sich zwischen den ökologischen Gruppen nur geringfügig.

Abbildung 11: Boxplots der AUC-Performancewerte für die bewerteten Artenverbreitungsmodelle (Zusammenfassung für alle modellierten Arten)

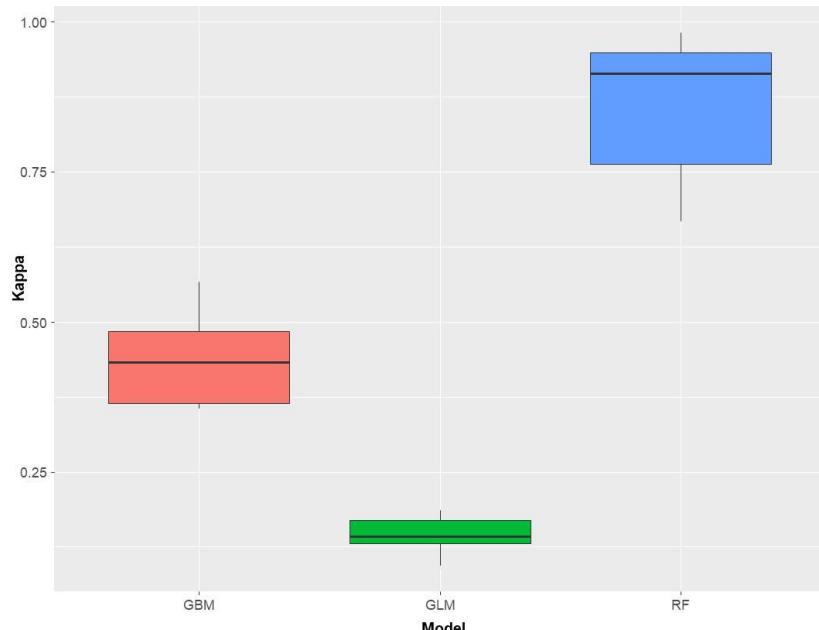
grün = GLM, rot = GBM, blau = RF; Kästchen umfassen das 25. bis 75. Quantil der Werte, Linien stellen das Minimum und Maximum dar, die horizontalen Linien den Mittelwert



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 12: Boxplots der Cohen-Kappa-Performanzwerte für die evaluierten Artenverteilungsmodelle (Zusammenfassung für alle modellierten Arten)

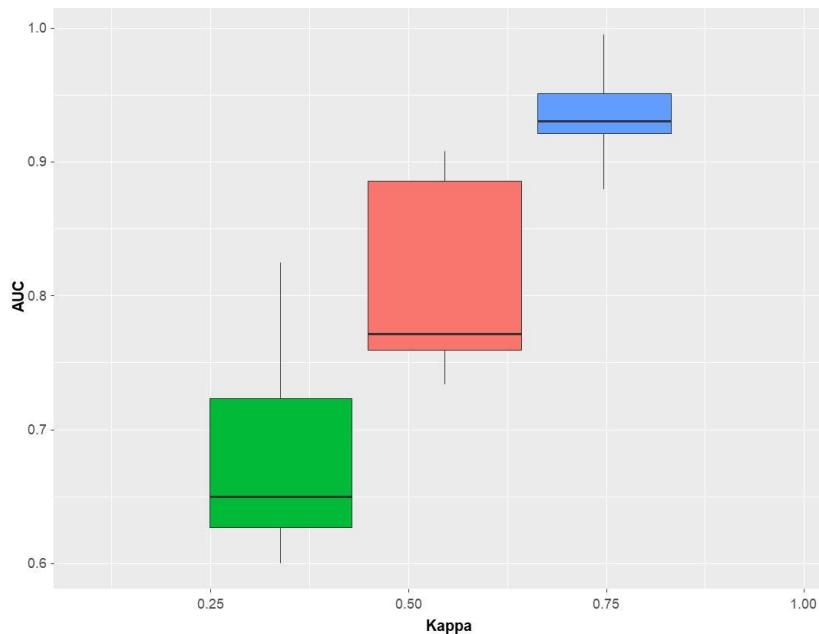
grün = GLM, rot = GBM, blau = RF; Kästchen umfassen das 25. bis 75. Quartil der Werte, Linien stellen das Minimum und Maximum dar, die horizontalen Linien den Mittelwert



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 13: Boxplots der kombinierten AUC- und Cohens Kappa-Werte für die bewerteten Artenverteilungsmodelle (Zusammenfassung für alle modellierten Arten)

grün = GLM, rot = GBM, blau = RF; Kästchen umfassen das 25. bis 75. Quartil der Werte, Linien stellen das Minimum und Maximum dar, die horizontalen Linien den Mittelwert



Quelle: Eigene Darstellung, Senckenberg Museum

Tabelle 7: Mittlere AUC_{ROC}- und Cohens Kappa-Werte der bewerteten Modelle aller ausgewählten Arten

Modell	<i>A. eiseni</i> ¹	<i>D. attemsi</i> ¹	<i>D. octaedra</i> ¹	<i>L. castaneus</i> ¹	<i>L. rubellus</i> ¹	<i>A. caliginosa</i> ²	<i>A. chlorotica</i> ²	<i>A. limicola</i> ²	<i>A. rosea</i> ²	<i>A. longa</i> ³	<i>L. badensis</i> ³	<i>L. terrestris</i> ³	Min	Max	Mittel
AUC															
GLM	0,825	0,706	0,654	0,624	0,680	0,621	0,628	0,775	0,645	0,601	0,796	0,646	0,601	0,825	0,683
GBM	0,908	0,890	0,807	0,762	0,763	0,772	0,734	0,884	0,752	0,742	0,896	0,771	0,734	0,908	0,807
RF	0,938	0,941	0,919	0,925	0,920	0,922	0,924	0,995	0,994	0,880	0,982	0,936	0,880	0,995	0,940
Cohens Kappa															
GLM	0,761	0,367	0,108	0,163	0,145	0,143	0,128	0,148	0,138	0,096	0,132	0,187	0,096	0,761	0,210
GBM	0,864	0,768	0,437	0,364	0,369	0,567	0,432	0,432	0,362	0,146	0,457	0,357	0,146	0,864	0,463
RF	0,903	0,923	0,974	0,778	0,876	0,938	0,718	0,669	0,941	0,357	0,982	0,974	0,357	0,982	0,836

Lebensformgruppen:

¹ epigäisch

² endogäisch

³ anöatisch

Die Leistung der Einzelarten-Modelle ist für alle Arten ähnlich gut, insbesondere was die AUC-Werte betrifft (Abbildung 14). Sowohl hohe AUC- als auch Kappa-Werte werden insbesondere für *A. eiseni*, *D. attemsi* und *L. badensis* erzielt, und die höchsten AUC-Werte für *A. limicola*, was auf die höchste Modellgenauigkeit für diese Arten hinweist. Interessanterweise repräsentieren

diese Arten diejenigen mit der geringsten Anzahl an Beobachtungen (N). Die niedrigsten AUC- und Kappa-Werte und damit die niedrigste Modellpräzision werden für *A. longa* beobachtet.

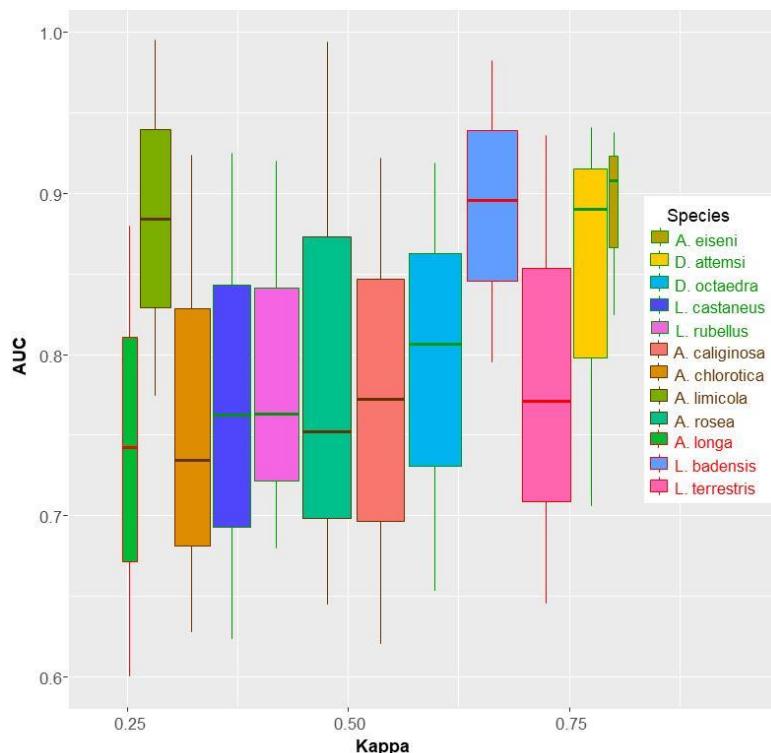
Daher haben wir für die endgültigen Modelle den Random Forest (RF)-Algorithmus zur Vorhersage der räumlichen Verteilung des Artenreichtums und der Gesamtabundanz von Regenwürmern sowie der Vorkommenswahrscheinlichkeiten von Arten in ganz Deutschland ausgewählt, basierend auf deren hoher Vorhersageleistung, die sich aus den aufgeführten Bewertungsmetriken ergibt.

3.2 Vorhergesagte räumliche Verteilung der Gesamtabundanz und des Artenreichtums der Regenwurmgesellschaften

Die vorhergesagte Gesamtabundanz der Regenwurmgesellschaften im Untersuchungsgebiet reicht von 10 bis maximal 800 Ind. m⁻², wobei die Modelle einen Durchschnitt von 350 Ind. m⁻² pro Standort vorhersagen (Abbildung 15). Die Vorhersagen zum Artenreichtum reichen von 1 bis 12 Arten, mit im Durchschnitt für ganz Deutschland durchschnittlich 3 Arten pro Standort (Abbildung 16). Höhere Gesamtabundanzen (> 400 Ind. m⁻²) werden für Grünland und Ackerland in Nordostdeutschland vorhergesagt (Abbildung 15). In diesen Regionen wird jedoch mit maximal 4 und durchschnittlich 2 Arten pro Standort ein relativ geringer Artenreichtum prognostiziert (Abbildung 16). Ein Vergleich mit prognostizierten Artenverteilungen (siehe unten) zeigt, dass in Nordostdeutschland vor allem epigäische Arten wie *D. octaedra* (aber teilweise

Abbildung 14: Kombinierte AUC- und Cohen-Kappa-Wert-Verteilung der einzelnen Arten aller Modelle zusammen (GBM, GLM, RF)

Kästchen umfassen das 25. bis 75. Quartil der Werte, Linien stellen das Minimum und Maximum dar, die horizontalen Linien den Mittelwert. Farbe des Artnamens (und Kästchenumrisses): grün epigäisch, braun endogäisch, rot anözisch lebende Art.



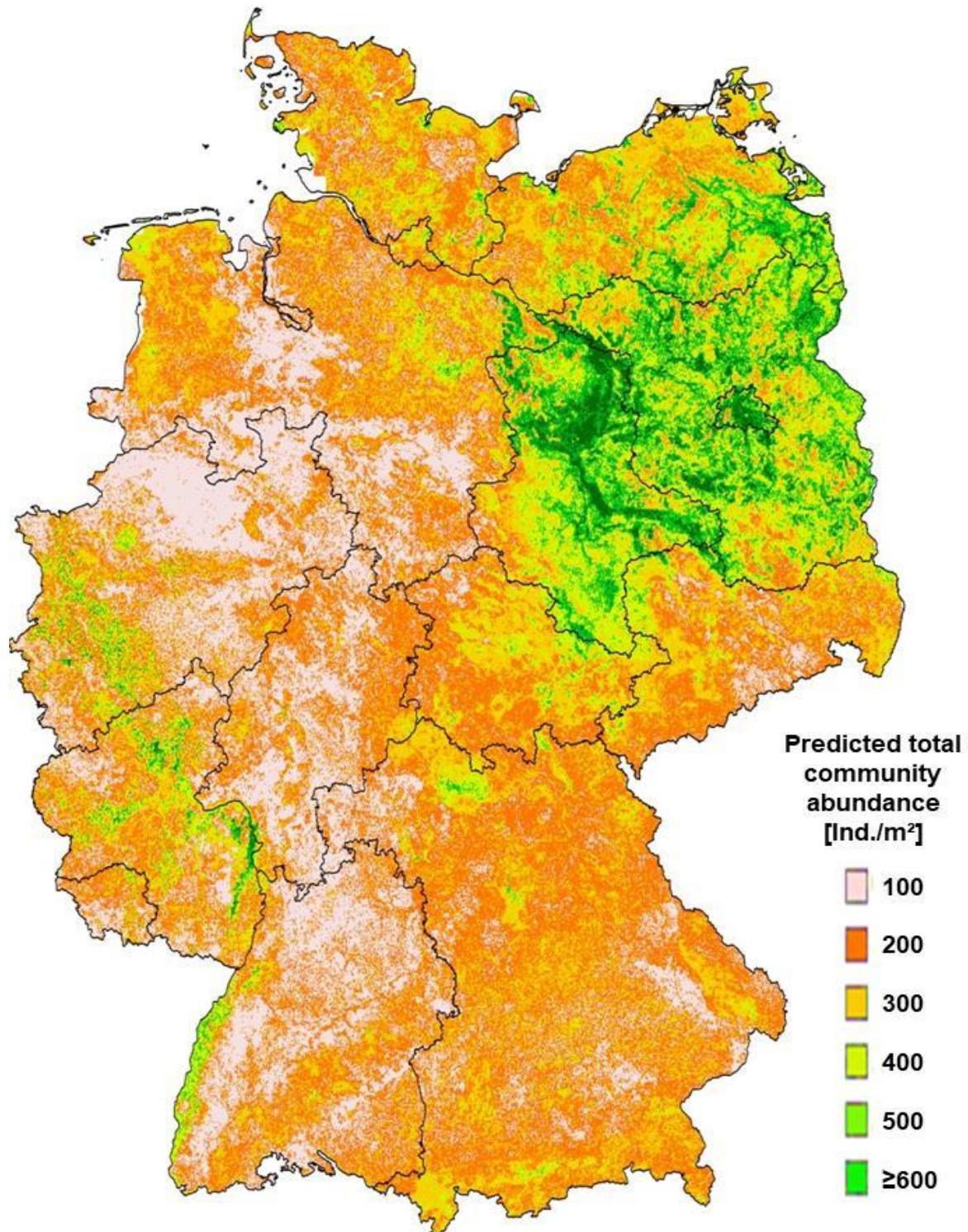
Quelle: Eigene Darstellung, Senckenberg Museum

auch *L. rubellus* und die endogäische *A. caliginosa*) für die prognostizierten hohen Gesamtabundanzen verantwortlich sind.

Die Überlagerung der Vorhersagen zum Artenreichtum und zur Gesamtabundanz ergibt eine Karte der Eignung für Regenwurmdiversität (Abbildung 17). Die Ergebnisse zeigen, dass vor allem das Rheintal, große Gebiete in Thüringen und die bayerischen Alpen Regionen mit potenziell hoher Regenwurmdiversität sind. Für diese Gebiete werden sowohl eine hohe Regenwurmgesamtabundanz als auch ein hoher Artenreichtum vorhergesagt, wobei die durchschnittliche Anzahl der vorkommenden Arten pro Standort 10 und die Gesamtabundanz 300 Ind. m⁻² erreicht. Die hohen Gesamtabundanzen in der Oberrheinebene werden zum Beispiel offenbar durch Arten weiter Verbreitung, wie *A. caliginosa*, *L. rubellus* und *A. rosea*, aber auch die hydrophile Art *L. limicola* verursacht, während die hohen Abundanzen der bayerischen Grünlandflächen auf eine höhere Artenzahl zurückgeht.

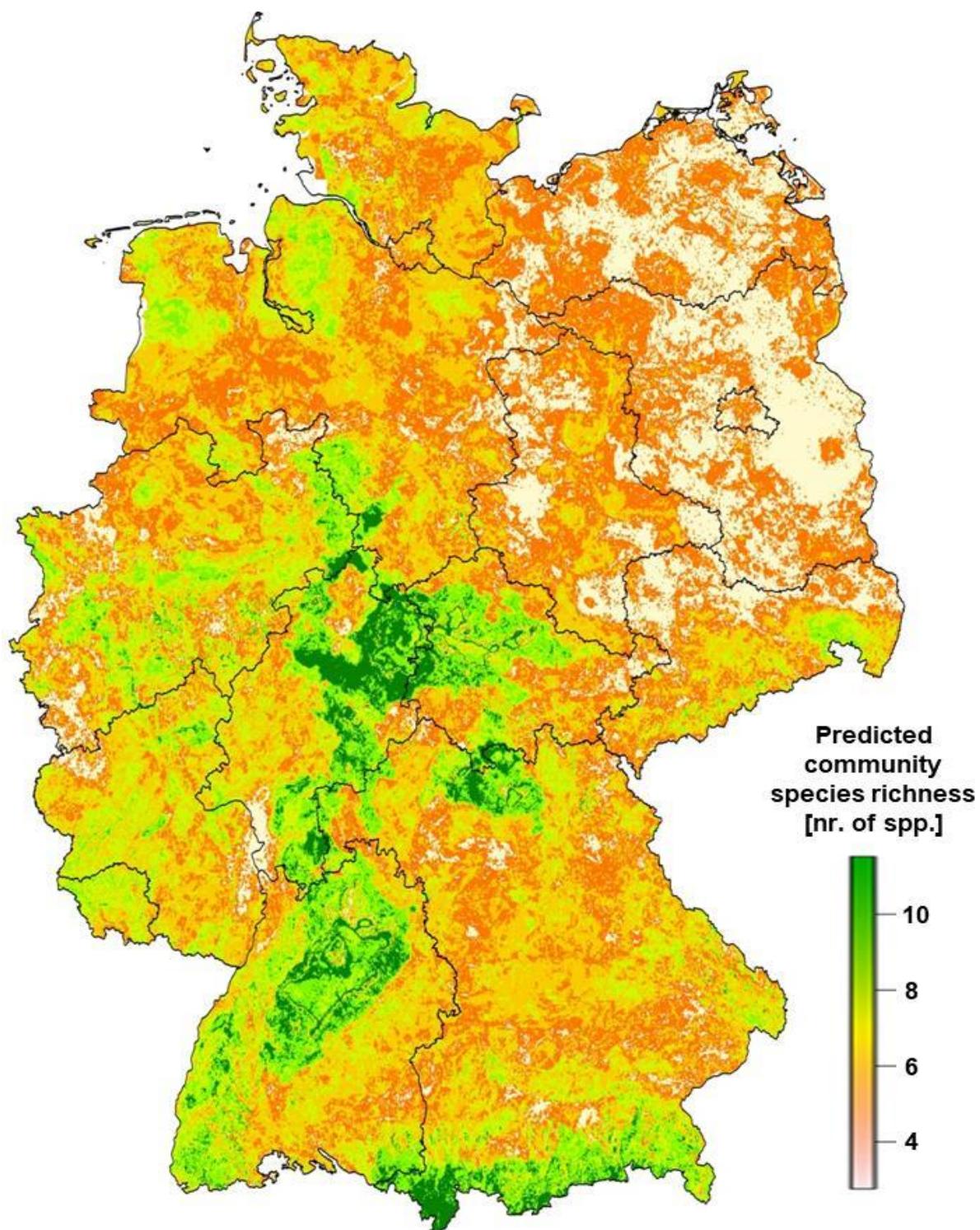
Abbildung 15: Gesamtabundanz der Regenwurmgesellschaften (RF-Modellvorhersagen)

Die Punkte stellen die ursprünglich beobachteten Daten dar, die zur Kalibrierung der Verbreitungsmodelle verwendet wurden (größere Punkte bedeuten eine höhere beobachtete Gesamtabundanz)



Quelle: Eigene Darstellung, Senckenberg Museum

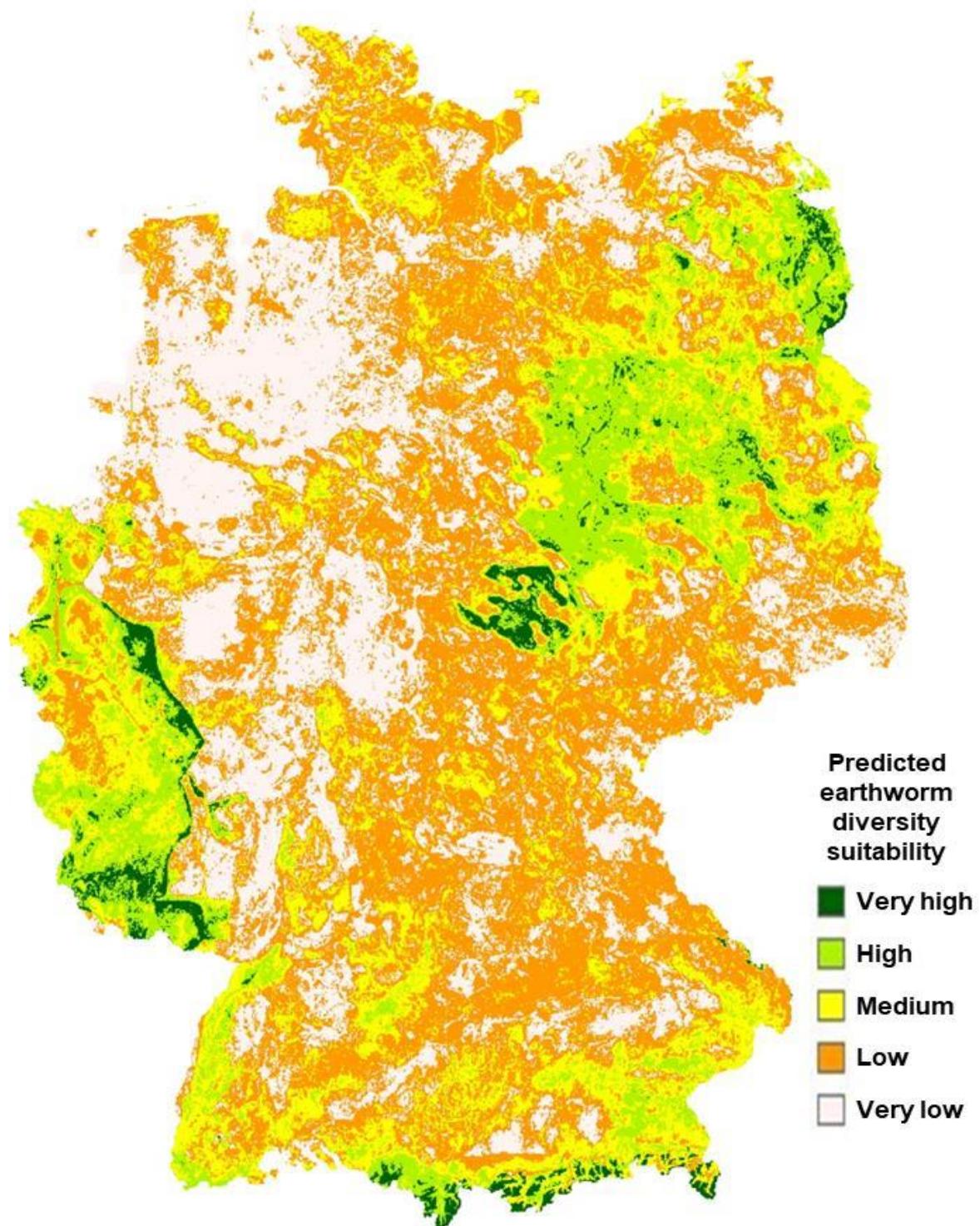
Abbildung 16: Artenreichtum von Regenwurmgemeinschaften (RF-Modellvorhersage)



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 17: Prognostizierte Erwartung zur Regenwurmdiversität

Überlagerung der RF-Modellvorhersagen des Artenreichtums und der Gesamtabundanz der Regenwurmgesellschaften. hellbraun = sehr niedrig, orange = niedrig, gelb = mittel, hellgrün = hoch, dunkelgrün = sehr hoch



Quelle: Eigene Darstellung, Senckenberg Museum

3.3 Räumliche Verbreitung ausgewählter Arten

Die modellierten Verbreitungsvorhersagen für einzelne Arten beruht auf Vorkommenswahrscheinlichkeiten, die sich auf einer Skala von 0 (nicht zu erwarten) bis 1 (höchste Vorkommenswahrscheinlichkeit) bewegen. Die aus den Modellergebnissen erzeugten Vorhersagekarten zeigen eine geographische Varianz der Verbreitung der ausgewählten Arten in Deutschland. Es lassen sich drei allgemeine Muster räumlicher Verbreitung in den Modellergebnissen unterscheiden: (1) Arten mit großer geographischer Verbreitung, (2) Arten mit mittlerer Verbreitung und (3) Arten mit geringer bis sehr geringer Verbreitung, einschließlich endemische Arten oder Arten, die auf sehr spezielle Habitate beschränkt sind (Abbildung 18 bis Abbildung 20; die Verbreitungsgebietsgrößen wurden auch als Gruppierungskriterium verwendet).

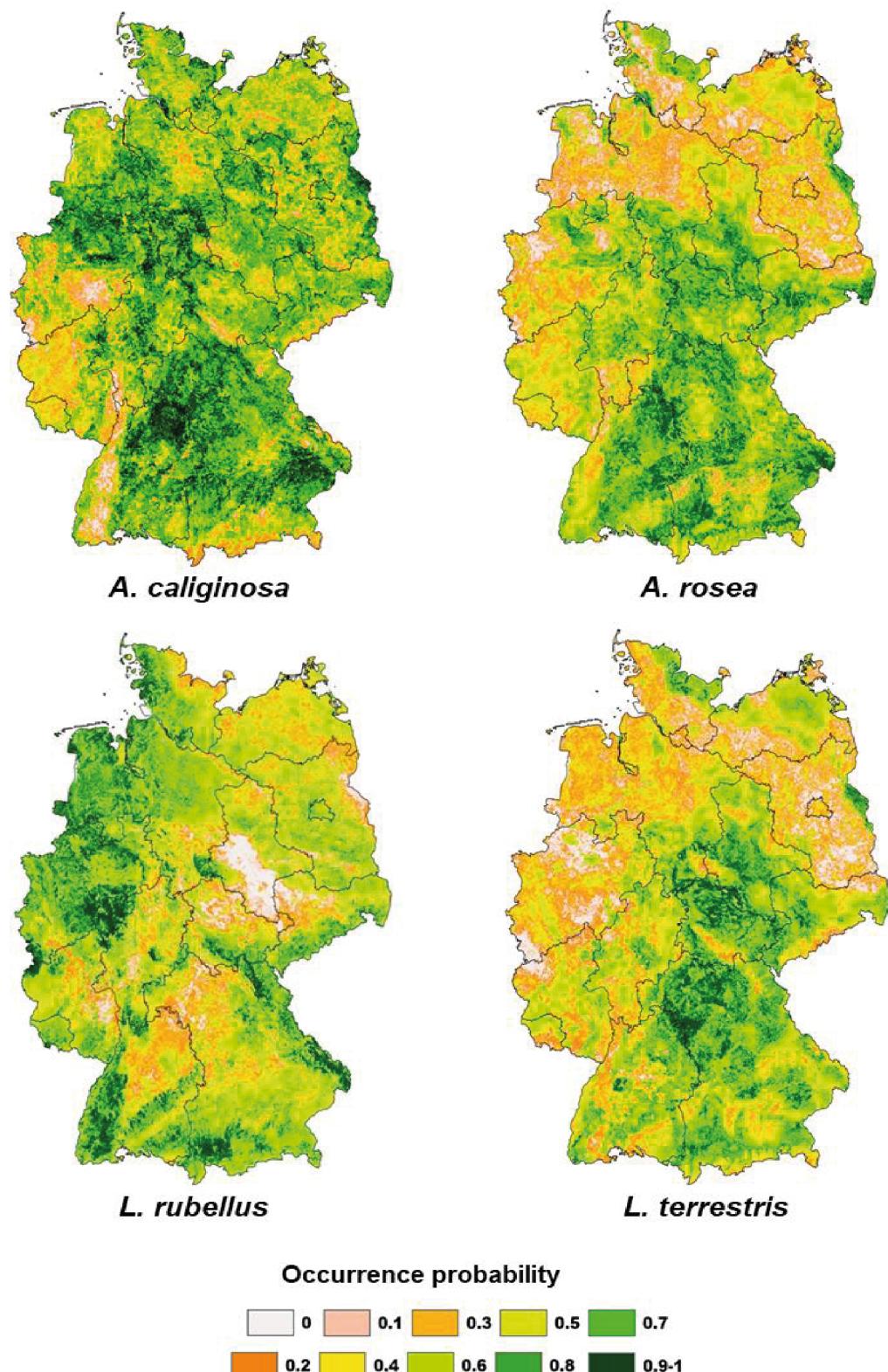
Die RF-Vorhersagen zeigen, dass die Arten großer geographischer Verbreitung in Deutschland (Abbildung 18) sich aus zwei endogäischen Arten (*A. caliginosa*, *A. rosea*) und je einer epigäischen und anözischen Art (*L. rubellus* bzw. *L. terrestris*) zusammensetzen, und dass *A. caliginosa* und *A. rosea* in vielen ökologischen Räumen Deutschlands weit verbreitet sind. Sie sind in den drei Hauptnutzungsformen Wald, Grünland und Acker verbreitet. Interessant ist die Prognose, dass *A. caliginosa* in Bergregionen im Südwesten bei Freiburg oder im Westen bei Siegen nur selten vorkommt. *A. rosea* ist nach den Vorhersagen auf Grünland in den Hochlagen der bayerischen Alpen weit verbreitet, aber in den nördlichen Bundesländern Niedersachsen, Brandenburg und Mecklenburg-Vorpommern nur schwach vertreten, mit Ausnahme einiger Häufungen von streifenförmigen Vorkommen im äußersten Osten Brandenburgs. Hingegen wird für *L. rubellus* in Wäldern und Ackerlandschaften in Westdeutschland, den Bundesländern Thüringen und Bayern im Vergleich zu den östlichen Regionen eine weite Verbreitung prognostiziert. Schließlich ist anzunehmen, dass *L. terrestris* am stärksten in zentralöstlichen und südlichen Regionen Deutschlands vorkommt, im Allgemeinen in den Gebieten, in denen für *L. rubellus* eine geringere Verbreitung prognostiziert wird.

Für die Arten mit mittlerer Verbreitung, dies sind die endogäische *A. cholorotica*, zwei epigäische Arten (*D. octaedra*, *L. castaneus*) und eine anözische Art (*A. longa*), wird eine hauptsächlich regional geclusternde Verbreitung prognostiziert, die insgesamt geringer ist als die der Arten mit großer Verbreitung (Abbildung 19). Für *A. cholorotica* werden Vorkommenszentren in Sachsen-Anhalt und Bayern vorhergesagt; *L. castaneus* in den meisten Bundesländern außer im Nordosten; *D. octaedra* vor allem in bergigen Waldgebieten aller Bundesländer sowie weiter verbreitet im nordöstlichen Tiefland.

Unter den Arten, denen ein eingeschränktes Verbreitungsgebiet vorhergesagt wird, befinden sich zwei epigäische Arten (*D. attemsi* und *A. eiseni*). Für diese Arten wird eine mittlere Verbreitung in wenigen Clustern in einigen Gebieten vorhergesagt (Abbildung 20, oben). Ein besonderes vorhergesagtes Verbreitungsmerkmal dieser Arten ist jedoch, dass sie im Allgemeinen nur eine geringe Wahrscheinlichkeit des Vorkommens aufweisen und in manchen Gebieten sogar fehlen. Beispielsweise wurde die Verbreitung von *A. eiseni* im Norden und im mittleren Bayern als sehr spärlich vorhergesagt, aber im mittleren Deutschland sowie im Schwarzwald als recht weit verbreitet. *D. attemsi* wird hauptsächlich in bewaldeten Gebieten der Mittelgebirgsgebiete, aber auch auf Äckern in Gebieten geringerer Meereshöhe prognostiziert.

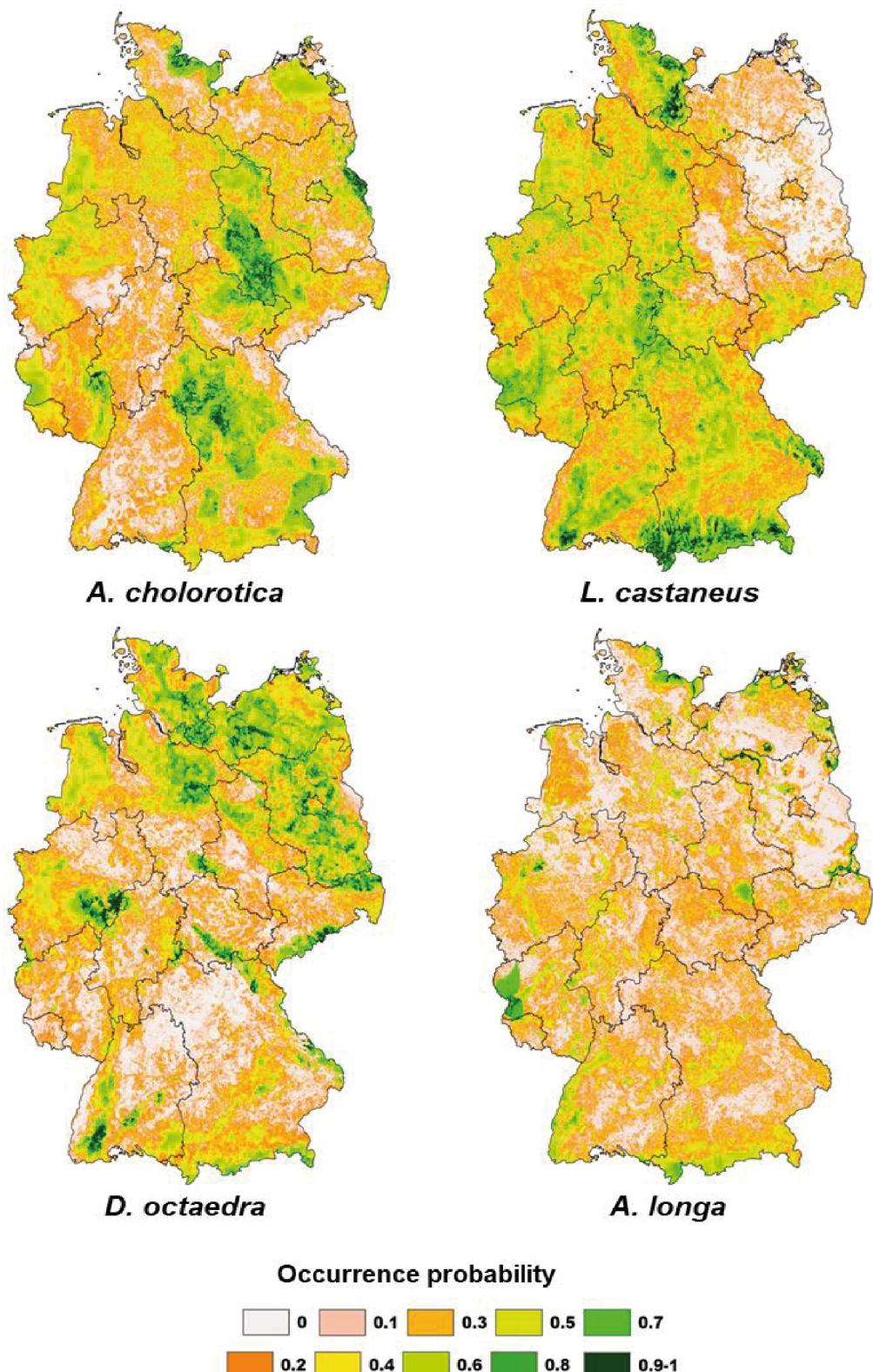
A. limicola und *L. badensis* stellen zwei Arten dar, die bekanntermaßen in besonderen Habitattypen vorkommen bzw. in Deutschland endemisch sind. Die Verbreitungsprognose weist *A. limicola* hauptsächlich entlang des Rheintals aus mit fleckenhaften Vorkommen in Feuchtgebieten in ganz Deutschland (Abbildung 20, unten links). Die Verbreitungsprognose für *L. badensis*

**Abbildung 18: Vorkommenswahrscheinlichkeiten von Arten weiter geographischer Verbreitung
(*A. caliginosa*, *A. rosea*, *L. rubellus*, *L. terrestris*) (RF-Modellprognosen)**



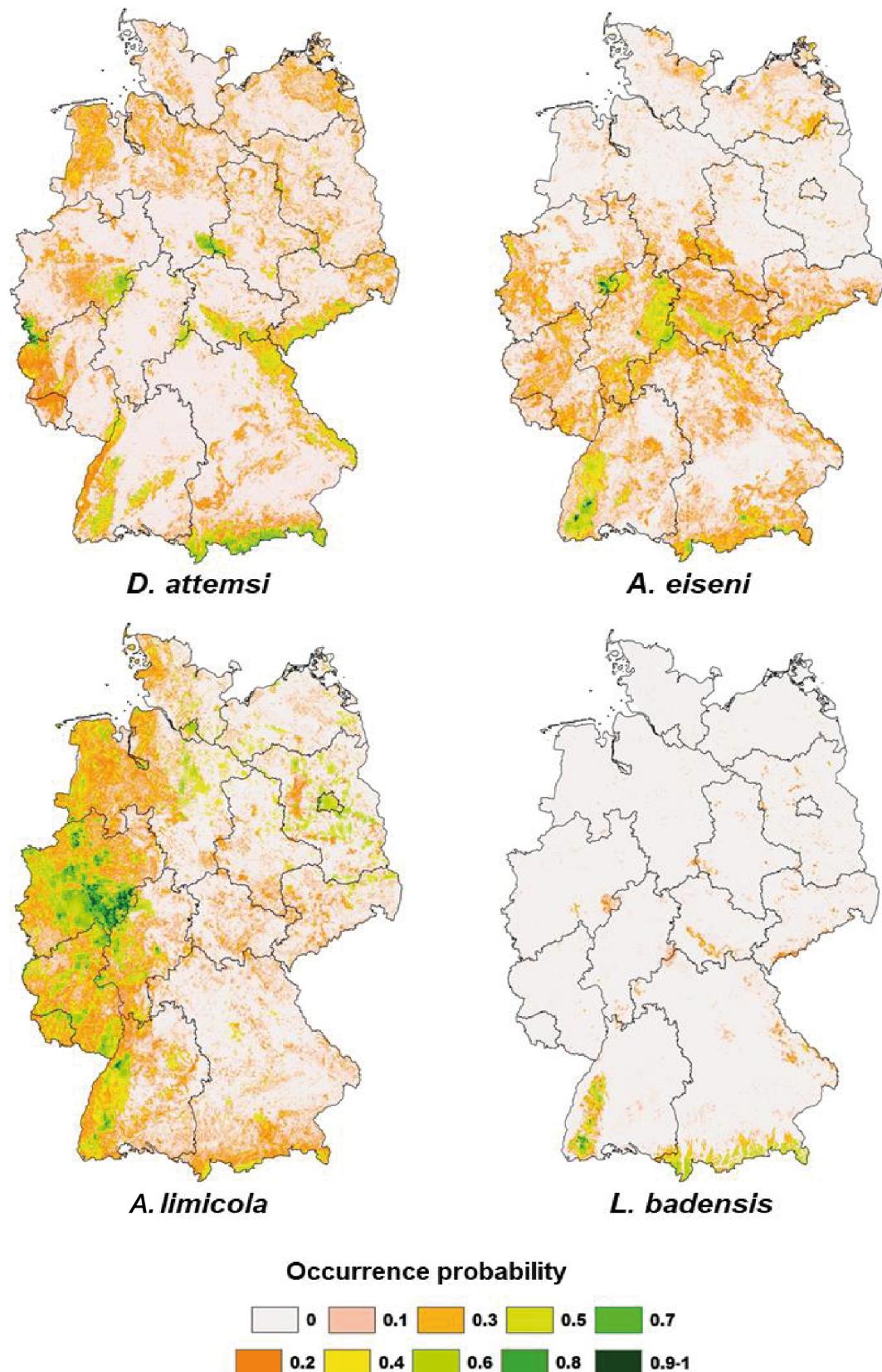
Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 19: Wahrscheinlichkeit des Vorkommens von Arten mittlerer geographischer Verbreitung (*A. chlorotica*, *D. octaedra*, *L. castaneus*, *A. longa*) (RF-Modellprognosen)



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 20: Wahrscheinlichkeit des Vorkommens von Arten mit eingeschränktem Verbreitungsgebiet (*D. attemsi*, *A. eiseni*) und Beispiele für Arten mit besonderem Habitat oder endemische Arten (*A. limicola* bzw. *L. badensis*) (RF-Modellprognosen)



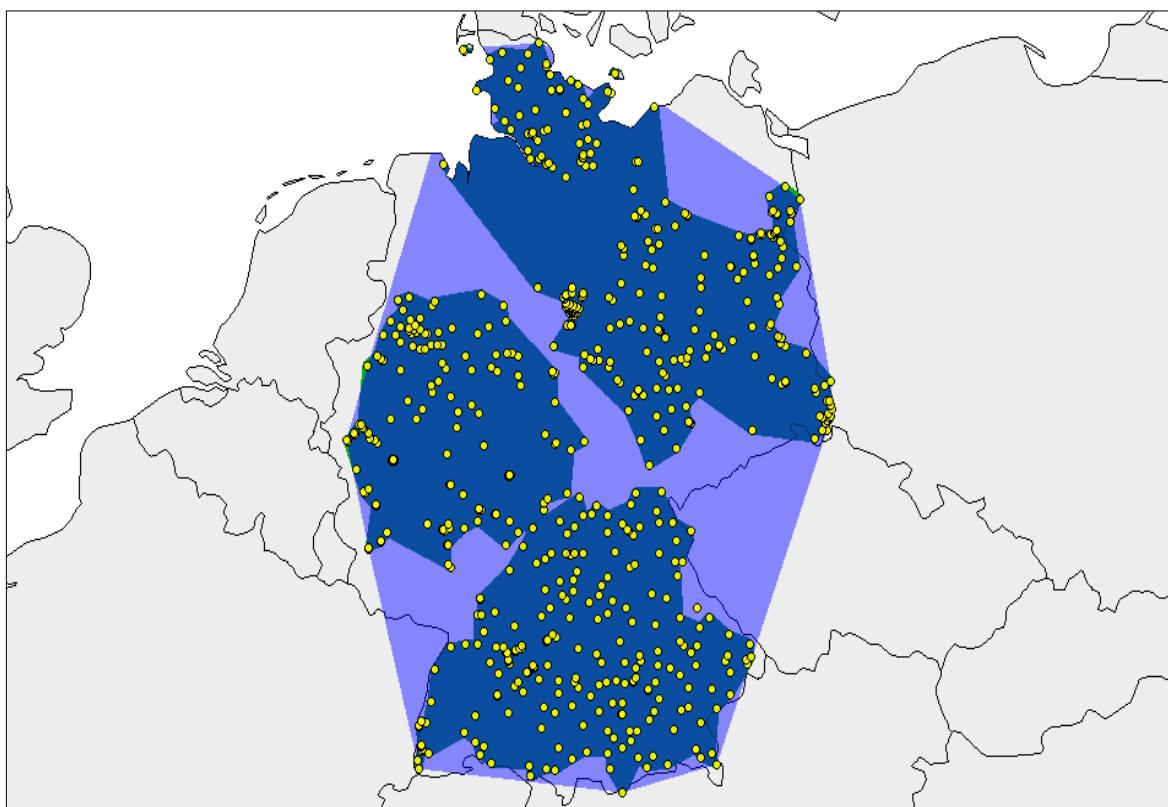
Quelle: Eigene Darstellung, Senckenberg Museum

(der „Badische Riesenregenwurm“) orientiert sich deutlich an der bekannten endemischen Verbreitung in der Schwarzwaldregion im Südwesten Deutschlands (Abbildung 20, unten rechts), aber die Modelle sagen auch eine mögliche Verbreitung in den bayerischen Alpen voraus .

3.4 Größe des geografischen Verbreitungsgebiets der ausgewählten Regenwurmarten

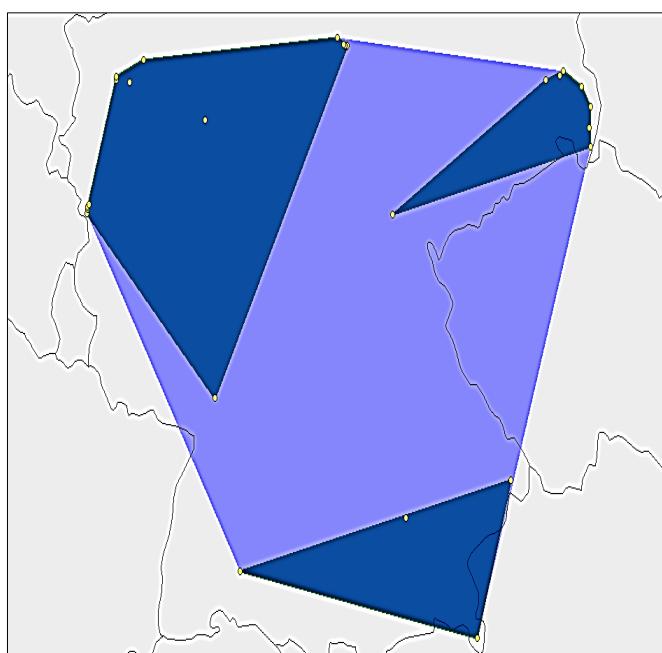
Die Größe des geografischen Verbreitungsgebiets in km^2 wurde unter Verwendung von EOO und AOO bestimmt, wie sie weithin verwendet und von der IUCN für die Einstufung von Arten in den Bedrohungsstatus empfohlen werden (IUCN 2022). Es gibt eine enge Übereinstimmung zwischen den modellierten Verbreitungskarten der Arten und der Größe ihrer geografischen Verbreitungsgebiete, da Arten mit weiter geographischer Verbreitung entsprechend größere EOOs und AOOs im Vergleich zu Arten mit geringer geographischer Verbreitung aufweisen. Abbildung 21 zeigt die EOO und die AOO von *A. caliginosa*, einer Art mit großem Verbreitungsgebiet, und Abbildung 22 jene von *D. attemsi*, einer Art mit eingeschränktem Verbreitungsgebiet.

Abbildung 21: Geografisches Verbreitungsgebiet von *A. caliginosa*: AOO (blau) und EOO (violett) mit tatsächlichen Fundpunkten (gelb) - (Die tatsächlichen Fundpunkte geben in erster Linie an, wo Aufzeichnungen verfügbar sind; in Gebieten ohne Fundpunkte sind ggf. tatsächliche Vorkommen nicht bekannt)



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 22: Geographisches Verbreitungsgebiet von *D. attemsi*: AOO (blau) und EOO (violett) mit tatsächlichen Fundpunkten (gelb) - (Die tatsächlichen Fundpunkte geben in erster Linie an, wo Aufzeichnungen verfügbar sind; in Gebieten ohne Fundpunkte sind ggf. tatsächliche Vorkommen nicht bekannt)



Quelle: Eigene Darstellung, Senckenberg Museum

Basierend auf den IUCN-Bewertungskategorien und den festgelegten Schwellenwerten für die Gefährdungskategorien (siehe Methoden) kann keine der analysierten Arten nur nach der EOO-Bewertung als bedroht angesehen werden, und fünf Arten sind weder nach der EOO-, noch nach der AOO-Bewertung bedroht (Tabelle 8). Basierend auf der Bewertung der AOO sind drei Arten als gefährdet und drei weitere Arten als sehr gefährdet einzustufen. Keine Art ist aufgrund von EOO und AOO als vom Aussterben bedroht einzustufen. Vier Arten mit ausgedehnten EOO > 360 000 km² und AOO > 2000 km² sind *A. caliginosa*, *A. rosea*, *L. terrestris* und *L. rubellus*. Die Vorkommensdaten für *L. badensis* (2 Präsenzpunkte) reichten nicht aus, um die Größe des geografischen Verbreitungsgebiets zu berechnen, da mindestens vier Vorkommensdaten erforderlich sind.

Tabelle 8: Bewertung der Gefährdung der Arten basierend auf den IUCN-Kategorie-B-Kriterien 1 und 2 (NA: Funddaten für die Berechnung der Bereichsgrößen nicht ausreichend)

Art	EOO (km ²) ¹	AOO (km ²) ¹	IUCN-Gefährdungskategorie
<i>A. caliginosa</i>	364624,4	2420	nicht gefährdet (EOO und AOO)
<i>A. rosea</i>	365299,4	2092	nicht gefährdet (EOO und AOO)
<i>L. terrestris</i>	365187,7	2084	nicht gefährdet (EOO und AOO)
<i>L. rubellus</i>	375796,7	2008	nicht gefährdet (EOO und AOO)
<i>A. chlorotica</i>	350969,0	2004	nicht gefährdet (EOO und AOO)
<i>L. castaneus</i>	355539,2	1040	nicht gefährdet (EOO); gefährdet (AOO)
<i>D. octaedra</i>	365051,9	1020	nicht gefährdet (EOO); gefährdet (AOO)

Art	EOO (km ²) ¹	AOO (km ²) ¹	IUCN-Gefährdungskategorie
<i>A. longa</i>	354399,3	572	nicht gefährdet (EOO); gefährdet (AOO)
<i>D. attemsi</i>	201500,8	108	nicht gefährdet (EOO); stark gefährdet (AOO)
<i>A. eiseni</i>	192725,0	100	nicht gefährdet (EOO); stark gefährdet (AOO)
<i>A. limicola</i>	262682,8	196	nicht gefährdet (EOO); stark gefährdet (AOO)
<i>L. badensis</i>	NA	NA	NA

¹ Schwellenwerte für EOO sind 100 km², 5000 km² und 20 000 km² für vom Aussterben bedrohte (CR), stark gefährdete (EN) bzw. gefährdete (VU) Arten, für AOO entsprechend 10 km², 500 km² and 2000 km².

3.5 Relativer Beitrag der Umweltvariablen zu den Modellvorhersagen und Verhalten von Gesellschaften und Arten gegenüber den Prädiktoren

Im Folgenden werden die relativen Beiträge der elf Umweltprädiktorvariablen zu den Verbreitungsmodellen sowie das Modellverhalten der Regenwurmgemeinschaften sowie jeder ausgewählten Art hinsichtlich dieser Parameter untersucht. Sowohl die relativen Beiträge als auch die Reaktionskurven zu diesen Variablen variieren und weisen bestimmte Muster auf.

3.5.1 Analysen auf Gemeinschaftsebene

3.5.1.1 Beiträge von Umweltprädiktorvariablen zu Modellergebnissen

Die Umweltvariablen, die am meisten zu den Vorhersagen der Gesamtabundanz von Regenwurmern beitragen, sind der durchschnittliche jährliche Niederschlag, die Bodenfeuchte, der Lehmgehalt und der Habitattyp, die zusammen über 40 % der Modellergebnisse ausmachen (Abbildung 23). Die verbleibenden Prädiktoren für die Gesamtabundanz tragen ziemlich gleichmäßig jeweils zwischen 5 und 10 % zu den Modellergebnissen bei, mit Ausnahme des Boden-pH-Werts, der interessanterweise nur zu weniger als 2 % zu den Vorhersagen beiträgt.

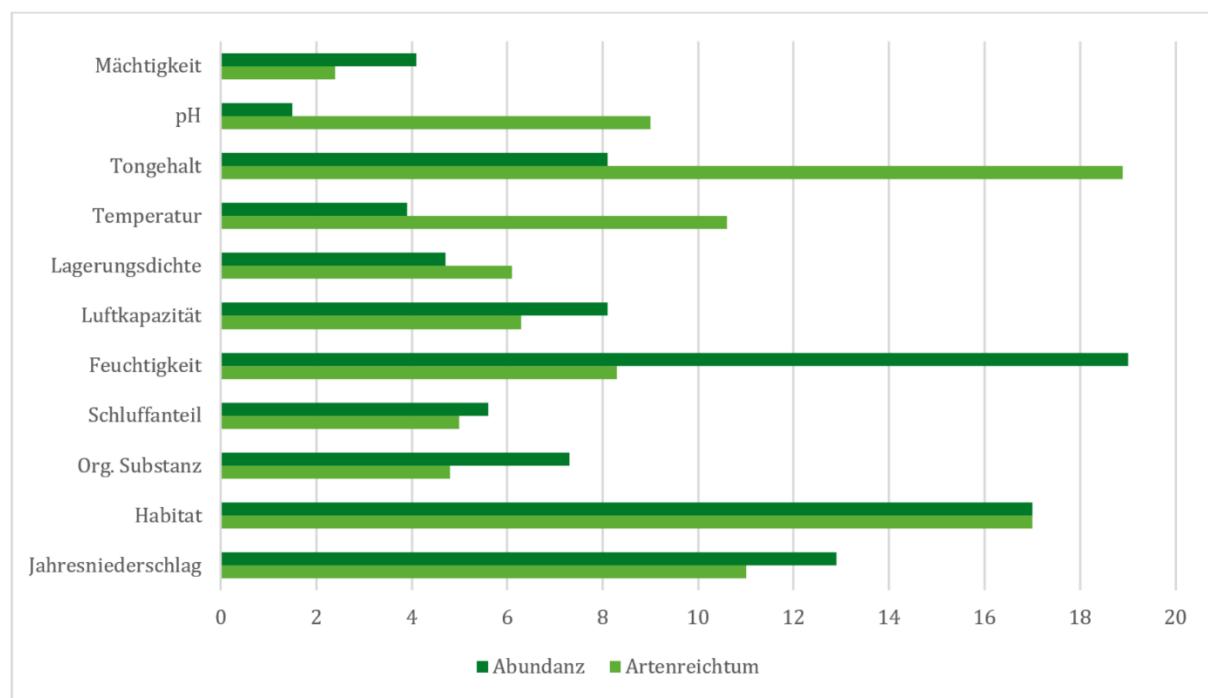
Der Beitrag der Umwelt zum vorhergesagten Artenreichtum wird vom Habitattyp und dem Tongehalt dominiert, die jeweils 17 bzw. 19 % beitragen (Abbildung 23). Der durchschnittliche jährliche Niederschlag und die Temperatur tragen jeweils mehr als 10 % zu den Vorhersagen des Artenreichtums bei. Im Gegensatz zu den Gesamtabundanzvorhersagen beeinflusst der pH-Wert auch die Artenreichtumsvorhersagen zu fast 10 %, während die übrigen Variablen recht gleichmäßig zwischen 4 und 8 % beitragen. Die Gründigkeit des Bodens spielt bei der Vorhersage des Artenreichtums die geringste Rolle.

3.5.1.2 Modellierter Response der Gesellschaften auf die Umweltprädiktorvariablen

Neben der Größe des Beitrags zu den Vorhersagen der Lebensgemeinschaften ist es von Bedeutung, wie und ab welchen Schwellenwerten Umweltparameter potenziell die Gesamtabundanz oder den Artenreichtum beeinflussen. Um dies abzuschätzen, wurden Responsekurven aus den Modellergebnissen abgeleitet.

Durchschnittliche Jahresniederschlagswerte von unter 500 mm a⁻¹ wirken sich erhöhend auf die Abundanzvorhersagen aus, über dieser Schwelle dagegen führen zu niedrigeren Abundanzprognosen (Abbildung 24). Die Jahresdurchschnittstemperatur hat keinen großen Einfluss auf die Abundanzvorhersagen, allerdings zeigen sich höhere modellierte Abundanzen oberhalb einem Schwellenwert von 10 °C. Interessanterweise haben diese Klimaparameter einen in gewisser Hinsicht gegenläufigen Effekt auf den vorhergesagten Artenreichtum, der

Abbildung 23: Beitrag (in %) der Umweltprädiktorvariablen zu Vorhersagen zur Gesamtabundanz und zum Artenreichtum der Regenwurmgemeinschaften



Quelle: Eigene Darstellung, Senckenberg Museum

Tabelle 9: Beitrag (in %) der Umweltprädiktorvariablen zu Vorhersagen zur Gesamtabundanz und zum Artenreichtum der Regenwurmgemeinschaften

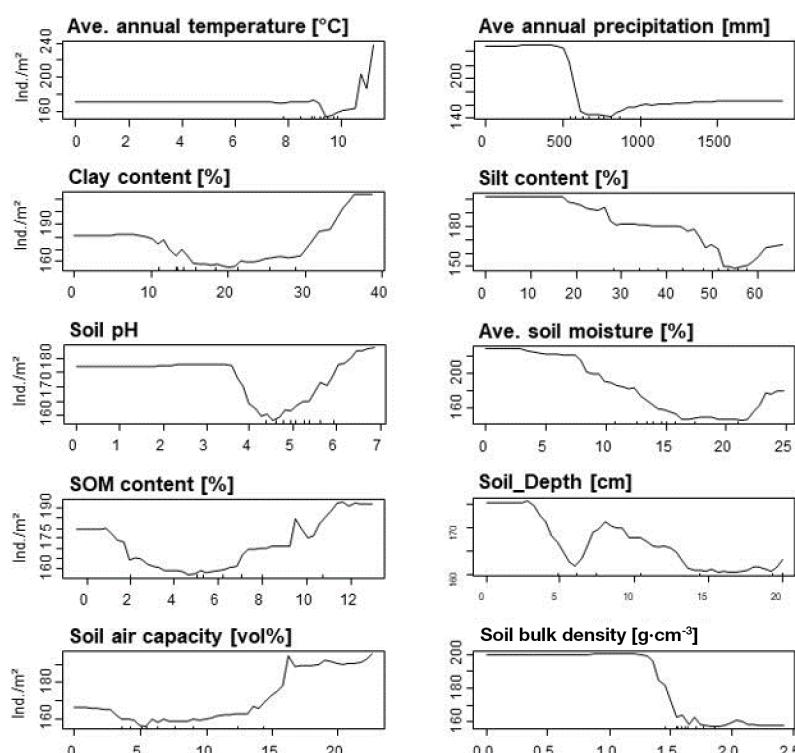
	Artenreichtum	Abundanz
Jahresniederschlag	11	12,9
Habitat	17	17
Org. Substanz	4,8	7,3
Schluffanteil	5	5,6
Feuchtigkeit	8,3	19
Luftkapazität	6,3	8,1
Lagerungsdichte	6,1	4,7
Temperatur	10,6	3,9
Tongehalt	18,9	8,1
pH	9	1,5
Mächtigkeit	2,4	4,1

vgl. Abbildung 23

über dem Schwellenwert von 500 mm a⁻¹ durchschnittlichem Jahresniederschlag zunimmt und bei Jahresschnittstemperaturen unter 6 °C ab (aber über 10 °C wiederum ansteigt) (Abbildung 25).

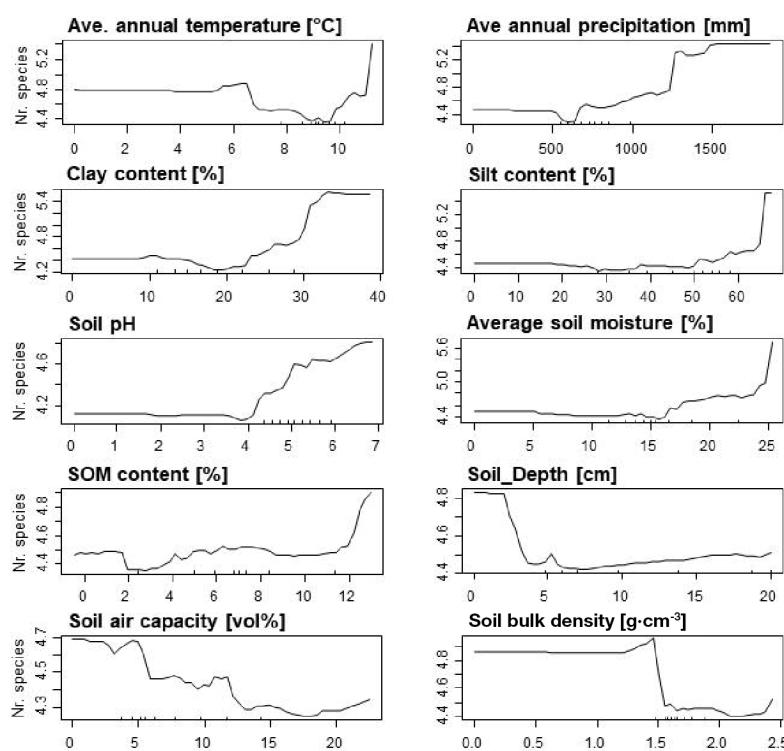
Hinsichtlich des Einflusses von Bodenparametern führt eine zunehmende durchschnittliche Bodenfeuchte (ab ca. 7 %) zu niedrigeren Gesamtabundanzen, beeinflusst jedoch die Vorhersagen für den Artenreichtum kaum (trotz einer starken vorhergesagten Zunahme in sehr feuchten Böden über ca. 20 % durchschnittlichem Wassergehalt). Die Bodenreaktion weist eine ähnliche Schwelle sowohl für die Gesamtabundanz- als auch für die Artenreichtumsprognosen auf. Während die Abundanzvorhersagen im gesamten pH-Bereich relativ konstant bleiben, nehmen sie bei etwa pH 4 deutlich ab, um oberhalb von pH 5 wieder zuzunehmen. Die Vorhersagen für den Artenreichtum bleiben dagegen unterhalb von pH 4 sehr niedrig und steigen oberhalb dieses Werts stark an. Obwohl die organische Substanz im Boden nur geringfügig zu den Modellergebnissen beiträgt, erhöht sie die Abundanzvorhersagen bei Gehalten über 6 %, während sie wenig Einfluss auf die Artenreichtumsvorhersagen zeitigt. Der Einfluss der Bodenart wird am Tongehalt am deutlichsten, wo die Vorhersagen sowohl für die Gesamtabundanz als auch für den Artenreichtum oberhalb 20 – 30 % Ton ansteigen. Der Schluffgehalt hat anscheinend wenig Einfluss auf den Artenreichtum, aber die Abundanzvorhersagen gehen über 30 % Schluffanteil stark zurück. Die Bodenstruktur beeinflusst auch die prognostizierten Regenwurmgemeinschaften. Bei einer Luftkapazität von über 15 % nehmen die Abundanzprognosen stark zu, die Artenreichtumsprognosen dagegen mit zunehmender Luftkapazität kontinuierlich ab. Bodenverdichtung wirkt sich offenbar negativ auf Regenwurmgemeinschaften aus, da sowohl die Gesamtabundanz- als auch die Artenreichtumsvorhersagen bei Lagerungsdichten über 1,5 g cm⁻³ abrupt und sehr deutlich abnehmen.

Abbildung 24: Responsekurven (partielle Abhängigkeit) der vorhergesagten Gesamtabundanz von Regenwürmern auf die ausgewählten Umweltprädiktorvariablen



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 25: Responsekurven (partielle Abhängigkeit) des vorhergesagten Artenreichtums von Regenwurmgesellschaften auf die ausgewählten Umweltprädiktorvariablen



Quelle: Eigene Darstellung, Senckenberg Museum

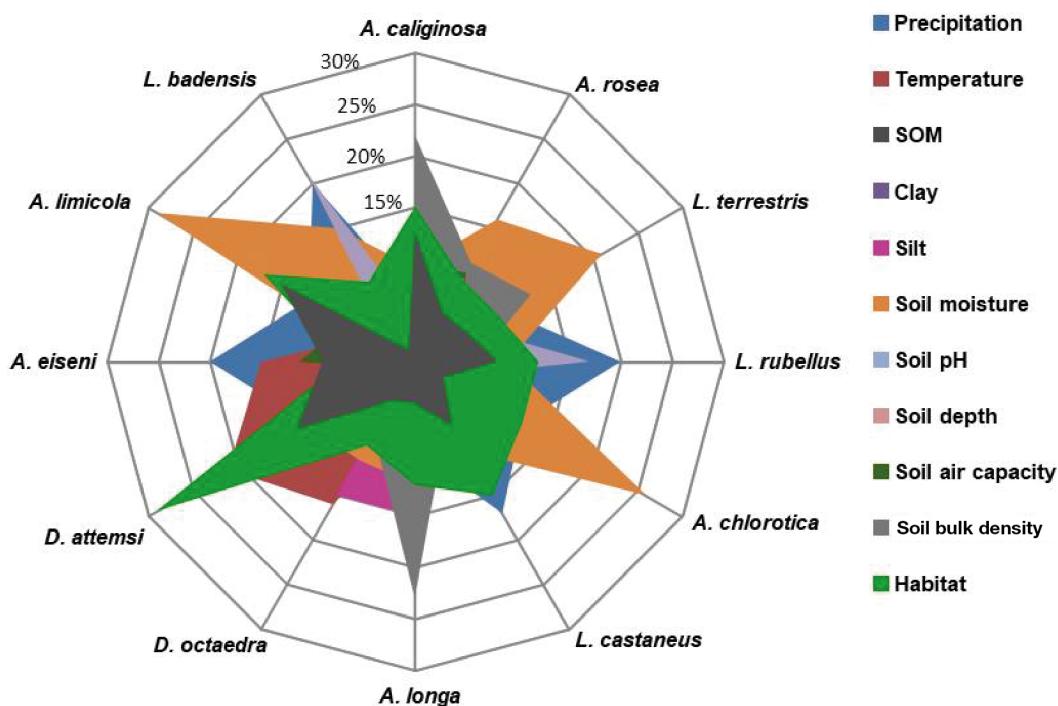
3.5.2 Artspezifische Analysen

3.5.2.1 Beiträge von Umweltprädiktorvariablen zu Modellergebnissen

Für die artspezifischen Modelle der Vorkommenswahrscheinlichkeit wurden vier Variablen als Haupttreiber der Verbreitung von Regenwurmarten in Deutschland identifiziert: Mittlerer Jahresniederschlag und die damit verbundene durchschnittliche Bodenfeuchte, Boden-pH-Wert und Habitattyp. Daneben sind für eine Reihe von Arten offenbar auch die Jahresdurchschnittstemperatur und der Gehalt des Bodens an organischer Substanz bedeutsam (Abbildung 26 und Tabelle 10).

Der mittlere Jahresniederschlag und die damit verbundene Bodenfeuchte machen bei vielen Arten 20 % oder mehr der Vorhersagen aus und erwiesen sich für die Modellierung der räumlichen Verbreitung von Arten wie *A. limicola*, *A. chlorotica* und *L. terrestris* als wesentliche Prädiktorvariablen (Abbildung 26 und Tabelle 10).

Abbildung 26: Relative Beiträge von Umweltprädiktorvariablen zu den modellierten Vorkommenswahrscheinlichkeiten ausgewählter Arten



Quelle: Eigene Darstellung, Senckenberg Museum

Tabelle 10: Relative Beiträge von Umweltprädiktorvariablen zu den modellierten Vorkommenswahrscheinlichkeiten von Arten

Art	NS	T	OS	Ton	U	Was- ser	pH	Grün- dig- keit	LK	LD	Habi- tat- typ
<i>A. caliginosa</i>	6	9,7	12,5	5,2	3,9	9,2	22	2,4	9,4	4,3	15,1
<i>A. chlorotica</i>	12	13	3,2	8,2	7	26	8	1,8	4,5	3	12
<i>A. eiseni</i>	20	15	9	10	4	3	7	2,4	11,2	3,1	7
<i>A. limicola</i>	12	5,2	15	3	2,3	29	2,6	4,2	7,7	2,2	17
<i>A. longa</i>	8,6	6,3	4,1	4,8	15	12	23	2,3	5,1	6,3	12
<i>A. rosea</i>	10	7,5	5,4	8,5	9,6	16	11	3	10	9,4	9,2
<i>D. attemsi</i>	12	21	13,2	4	1,3	12	9	3,4	0,9	4,5	29
<i>D. octaedra</i>	8	16	4,4	6,5	15	11	8	4,5	7,9	8	9,4
<i>L. badensis</i>	20	0	1,2	0	2	15	0	0	3	21	9
<i>L. castaneus</i>	17	5,2	7,2	9	6,6	10	6,4	4,5	11	6,5	15
<i>L. rubellus</i>	20	10	8	5	4	9	6	2,2	7	17	12
<i>L. terrestris</i>	10	9	5,7	5,1	11	21	13	2,5	5	6,5	9,2

NS Mittlerer Jahresniederschlag, T Jahresdurchschnittstemperatur; Gehalt des Bodens an OS organischer Substanz, Ton Ton, U Schluff; Wasser mittlerer Bodenwassergehalt, pH Bodenreaktion, Gründigkeit des Bodens, LK Luftkapazität, LD effektive Lagerungsdichte

Der Habitattyp wurde als nächstwichtiger Umweltprädiktor für die Wahrscheinlichkeit des Vorkommens von Arten identifiziert. Bei einem durchschnittlichen Beitrag von 13 % reichen die Beiträge zu einzelnen Arten von 9,2 bis 29 % (Tabelle 10).

Der Beitrag der Bodenreaktion zu den Vorkommensvorhersagen ist bei allen Arten moderat und reicht von 0 bis 13 % mit einem Durchschnitt von 10 %. Ausnahmen sind *A. longa* und *A. caliginosa*, deren Verbreitung stark vom pH-Wert des Bodens abhängt und der einen außergewöhnlich hohen Beitrag zur Auftretenswahrscheinlichkeit von > 20 % leistet (Tabelle 10).

Für den Gehalt an organischer Substanz im Boden ist ein relativ höherer prozentualer Beitrag zu den Vorkommensvorhersagen für epigäische Arten wie *D. attemsi* (13,2 %), *A. eiseni* (9 %) und *L. castaneus* (7,2 %) festzustellen, im Vergleich zu viel niedrigeren Vorhersagebeiträgen von weniger als 6 % für endogäische und anözische Arten wie *A. rosea*, *A. chlorotica*, *A. longa* und *L. terrestris*. Ausnahmen sind hier die beiden endogen lebenden Arten *A. caliginosa* und *A. limicola*, deren Auftretenswahrscheinlichkeiten stark vom Gehalt an organischer Substanz im Boden abhängt (12,5 bzw. 15 %).

Die Bodenart (basierend auf den Ton- und Schluffgehalt als Proxies) hat vergleichsweise weniger Einfluss auf das Vorkommen der Arten, mit Ausnahme von *A. longa* und *D. octaedra*, wo der Schluffgehalt 15 % zu den Vorkommensvorhersagen beiträgt. Das Vorkommen der Arten *D. attemsi*, *A. eiseni*, *A. chlorotica* und *L. rubellus* hängt auch stark vom Klima (mittlerer Jahresniederschlag und Jahresdurchschnittstemperatur) ab mit einem entsprechend großen Beitrag dieser Variablen (meist > 21 %) zu den Vorkommensvorhersagen. Auch die Auftretenswahrscheinlichkeit von *L. badensis* hängt in hohem Maße (20 %) mit dem mittleren Jahresniederschlag zusammen. Die Gründigkeit trägt am wenigsten zu den Modellergebnissen bei und beeinflusst die Vorkommenswahrscheinlichkeiten der meisten Arten nur mit einem Beitrag von etwa 2–5 %.

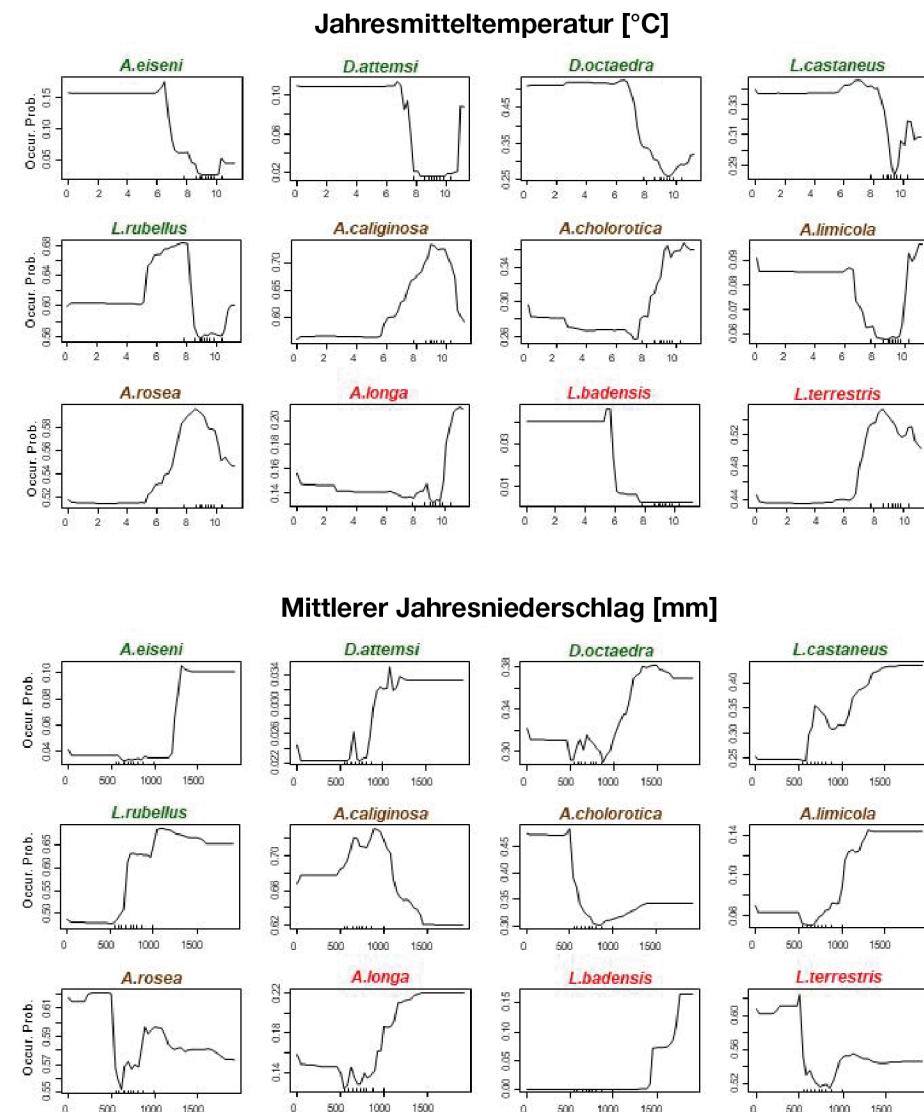
3.5.2.2 Modellierter Response der Arten auf Umweltprädiktorvariablen

Der modellierte Response der Arten auf quantitative Umweltvariablen zeigt drei Muster: einen positiven Zusammenhang, einen negativen Zusammenhang oder unregelmäßige oder bimodale Responsemuster entlang eines Gradienten zunehmender Klima- oder Bodenparameterwerte.

Die Klimaparameter (Jahresdurchschnittstemperatur und mittlerer Jahresniederschlag) und die damit verbundene durchschnittliche Bodenfeuchte zeigen ähnliche Einflussmuster auf die Vorkommensvorhersagen. Sie unterscheiden sich aber zwischen epigäischen und endogäischen Arten (Abbildung 27, Abbildung 31). Die Vorkommenswahrscheinlichkeiten epigäischer Arten nehmen im Allgemeinen mit steigender Temperatur ab (normalerweise bei Jahresmitteltemperaturen über 6 – 8 °C) und nehmen umgekehrt mit steigenden mittleren Jahresniederschlägen und Bodenfeuchten zu. Die Schwellenwerte, bei denen die Auftretenswahrscheinlichkeiten ansteigen, sind artspezifisch: die Wahrscheinlichkeiten von z. B. *D. attemsi*, *L. castaneus* und *L. rubellus* steigen bei Niederschlagsmengen von über 500 mm a⁻¹ und die von *A. eiseni* und *D. octaedra* von über ca. 1000 mm a⁻¹ an (Abbildung 27). Erhöhte Vorkommenswahrscheinlichkeiten epigäischer Arten wurden für unterschiedliche durchschnittliche Bodenfeuchte festgestellt (Abbildung 31), mit Ausnahme der Art *A. eiseni*, deren Vorkommensvorhersagen bei durchschnittlichen Wassergehalten im Boden oberhalb von ca. 6 % abnehmen. Im Gegensatz dazu nehmen Wahrscheinlichkeiten des Vorkommens endogäischer Arten mit höherer Jahresmitteltemperatur zu und umgekehrt mit niedrigerer mittlerer Jahresniederschlagsmenge und Bodenfeuchte ab. Auch hier sind die Schwellenwerte, bei denen sich die Vorhersagen des Auftretens ändern, artspezifisch (Abbildung 27, Abbildung 31). Eine Ausnahme von diesem Muster stellt die endogäische *A. limicola* dar, deren Vorkommenswahrscheinlichkeiten mit niedrigerer Jahresmitteltemperatur (unter ca. 7 °C) und höherem Jahresniederschlag (ca. 800 mm a⁻¹) und mittlerer Bodenfeuchte (> ca. 18 % H₂O) zunimmt. Anözische Arten zeigten kein allgemeines Muster, wobei die Vorhersagen für *L.*

terrestris zunehmen und die Vorhersagen für *L. badensis* abnehmen, wenn die mittlere Jahres-temperatur höher und die mittlere jährliche Niederschlagsmenge und Bodenfeuchte niedriger ausfallen. Die Vorkommenswahrscheinlichkeiten von *A. longa* steigen sowohl mit zunehmender Temperatur als auch mit zunehmendem Niederschlag und zunehmender Bodenfeuchte.

Abbildung 27: Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten der Arten auf die Variablen mittlerer Jahresniederschlag und Temperatur (= partielle Abhängigkeiten der Umweltprädiktoren von den prognostizierten Vorkommenswahrscheinlichkeitsverteilungen). Grün markierte Artennamen gelten im Allgemeinen als epigäisch, braun als endogäisch und rot als anözisch

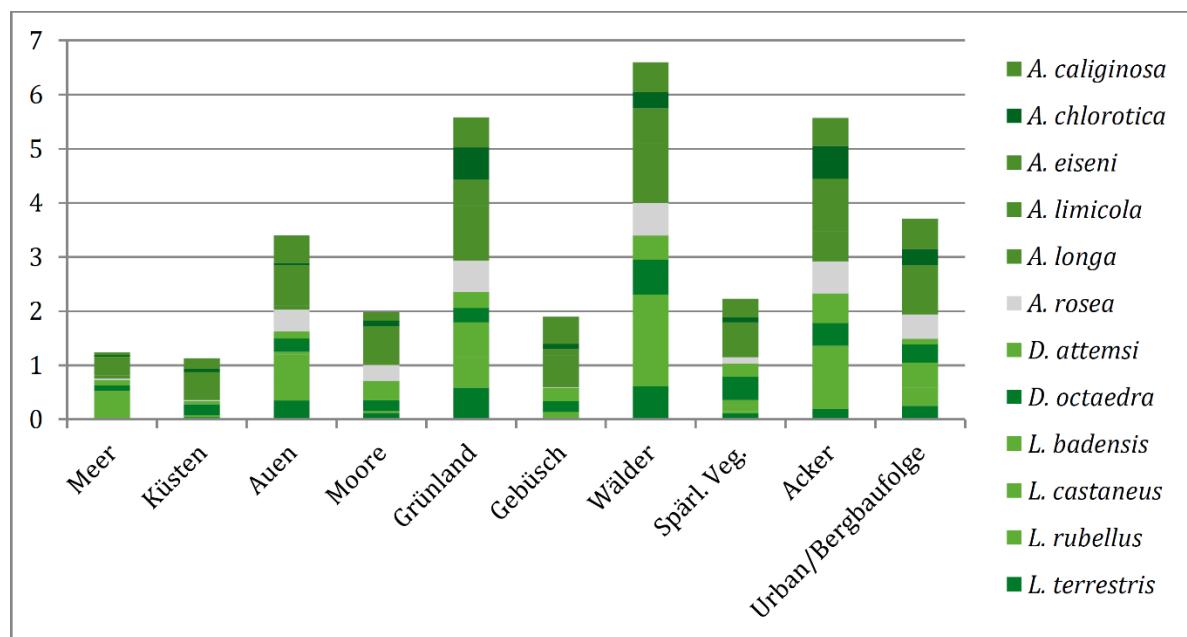


Quelle: Eigene Darstellung, Senckenberg Museum

Für den Habitattyp als kategoriale Variable (im Gegensatz zu den anderen, kontinuierlichen numerischen Variablen) konnten aus den Modellergebnissen keine Responsekurven für die einzelnen Arten abgeleitet werden. Um die vorhergesagten Habitatpräferenzen der Arten zu bestimmen, wurden daher die modellierten Vorkommenswahrscheinlichkeiten in jedem Habitattyp verwendet. Die Wahrscheinlichkeiten des Vorkommens der Arten wurden für zehn Haupthabitattypen modelliert (siehe Kap. 3). Dabei zeigen sich die größten Vorkommenswahrscheinlich-

keiten (die auf Habitatpräferenzen hindeuten) hauptsächlich für die Habitattypen Wald (kombiniert Laub- und Nadelwald), Grünland und Acker. Die Habitate der Nord- und Ostseeinseln, der Küsten sowie der Hoch- und Niedermoore und Gebüsche sind die am wenigsten bevorzugten (Abbildung 28 und Abbildung 29). Da die maximale vorhergesagte Wahrscheinlichkeit des Auftretens in einem Habitat für alle Arten 0,65 betrug, wurde eine vorhergesagte Wahrscheinlichkeit von 0,25 als Schwellenwert für die Angabe der Habitatpräferenz einer Art festgelegt (Abbildung 29). Die Vorkommensmodelle identifizierten einige Arten mit möglicherweise besonderen oder Haupthabitatpräferenzen, wie *D. octaedra* (z. B. Abbildung 30), *L. badensis* und *A. eiseni* im Wald oder *A. chlorotica* und *D. attemsi* in landwirtschaftlich geprägten Habitaten (Acker und Grünland) (Abbildung 29, Tabelle 11). *A. limicola* wird insbesondere für Feuchtgebiete wie Auen und Moore mit hoher Wahrscheinlichkeit prognostiziert. Interessant waren die prognostizierten hohen Vorkommenswahrscheinlichkeiten für eine Reihe von Arten in städtischen Gärten und Parks sowie in Bergbaufolgelandschaften (Abbildung 31). Für die meisten Arten wurde vorhergesagt, dass sie in allen Lebensräumen zumindest mit einer Wahrscheinlichkeit von wenigen Prozent vorkommen. Nur für *L. castaneus* wird für die Küstengebiete und für *L. badensis* für Gebüsche ein völliges Fehlen vorhergesagt. Abgesehen von Wäldern (50 % Vorkommenswahrscheinlichkeit), Grünland (30 %) und Ackerland (20 %) soll *L. badensis* in den meisten anderen Habitattypen selten vorkommen (weniger als 5 % Wahrscheinlichkeit). Für Arten wie *A. chlorotica*, *A. longa* oder *D. attemsi* wurde vorhergesagt, dass sie in marinen, küstennahen oder Auenhabitaten generell nur selten vorkommen.

Abbildung 28: Kumulative modellierte Vorkommenswahrscheinlichkeiten in den Habitattypen, für die am häufigsten ein hohes Vorkommen von Regenwurmarten vorhergesagt wird

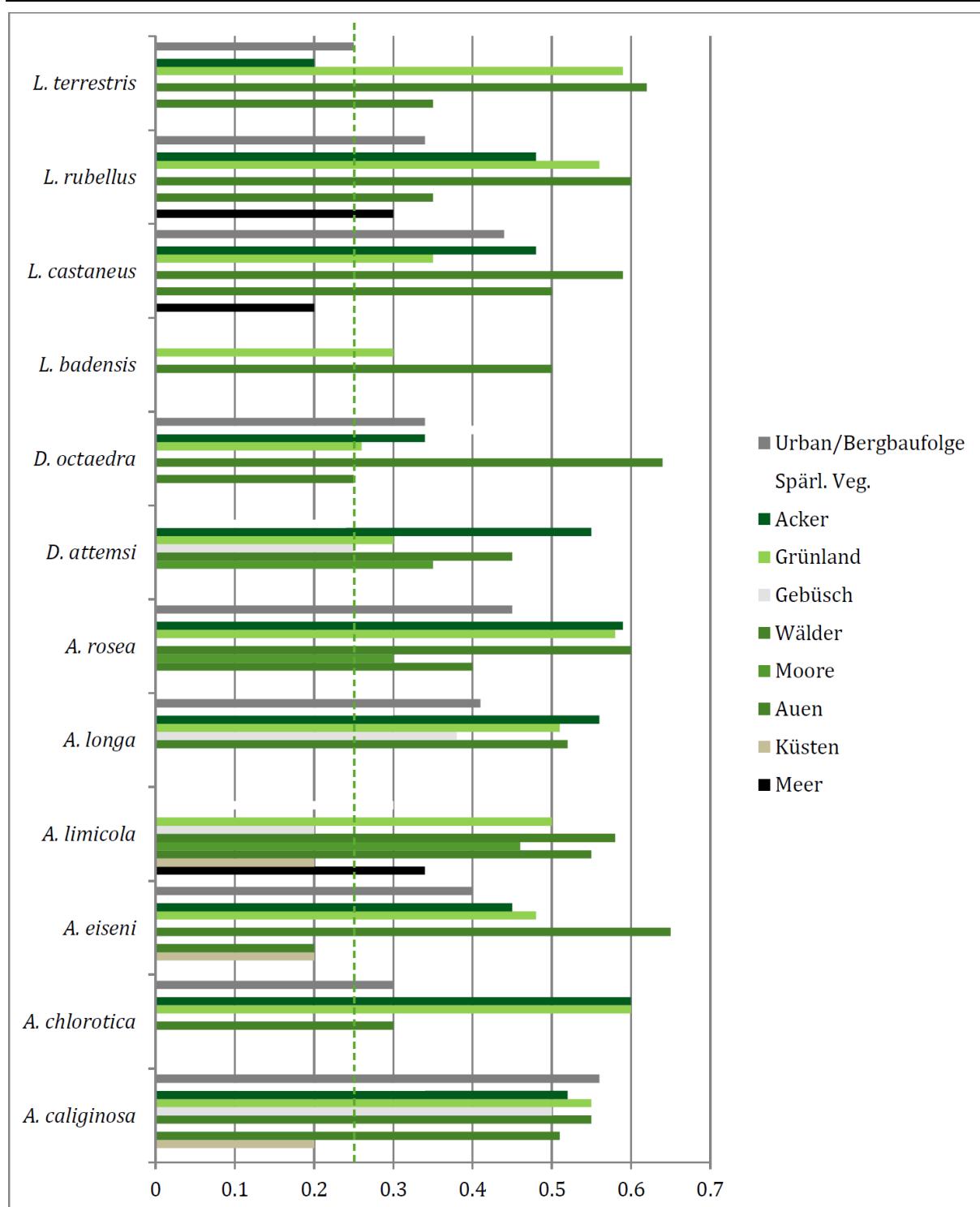


* Meereslebensräume sind hauptsächlich Inseln in der Ost- und Nordsee, die stark von maritimen Bedingungen beeinflusst werden

Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 29: Prognostizierte Habitatvorkommenswahrscheinlichkeiten der Arten über alle Habitatklassen.

Vertikale gepunktete Linie: Vorkommenswahrscheinlichkeit 0,25 als minimaler Vorfahrseschwellenwert, der die Habitatpräferenzen der Arten angibt.



* Meereslebensräume sind hauptsächlich Inseln in der Ost- und Nordsee, die stark von maritimen Bedingungen beeinflusst werden

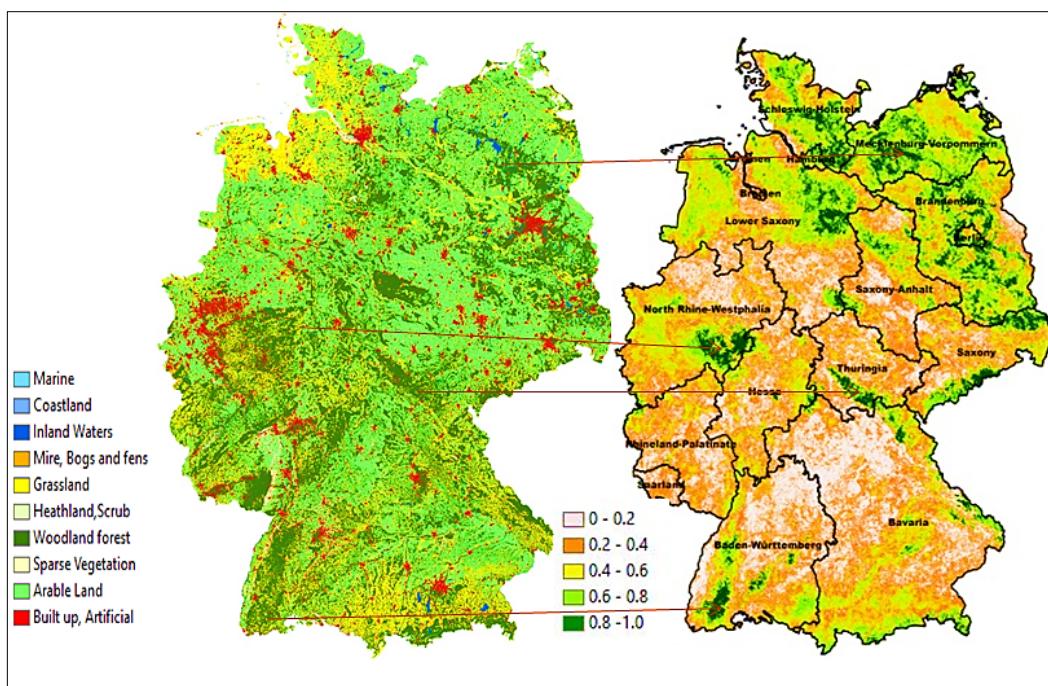
Quelle: Eigene Darstellung, Senckenberg Museum

Tabelle 11: Potenzielle Habitatpräferenzen der analysierten Arten (nach Prognose der Vorkommenswahrscheinlichkeitsmodelle)

Primärhabitare sind durch eine prognostizierte Auftrittswahrscheinlichkeit > 50 % bestimmt; Sekundärhabitare mit prognostizierten Wahrscheinlichkeiten zwischen 40 und 50 %. Habitate geordnet nach (abnehmender) Vorkommenswahrscheinlichkeit.

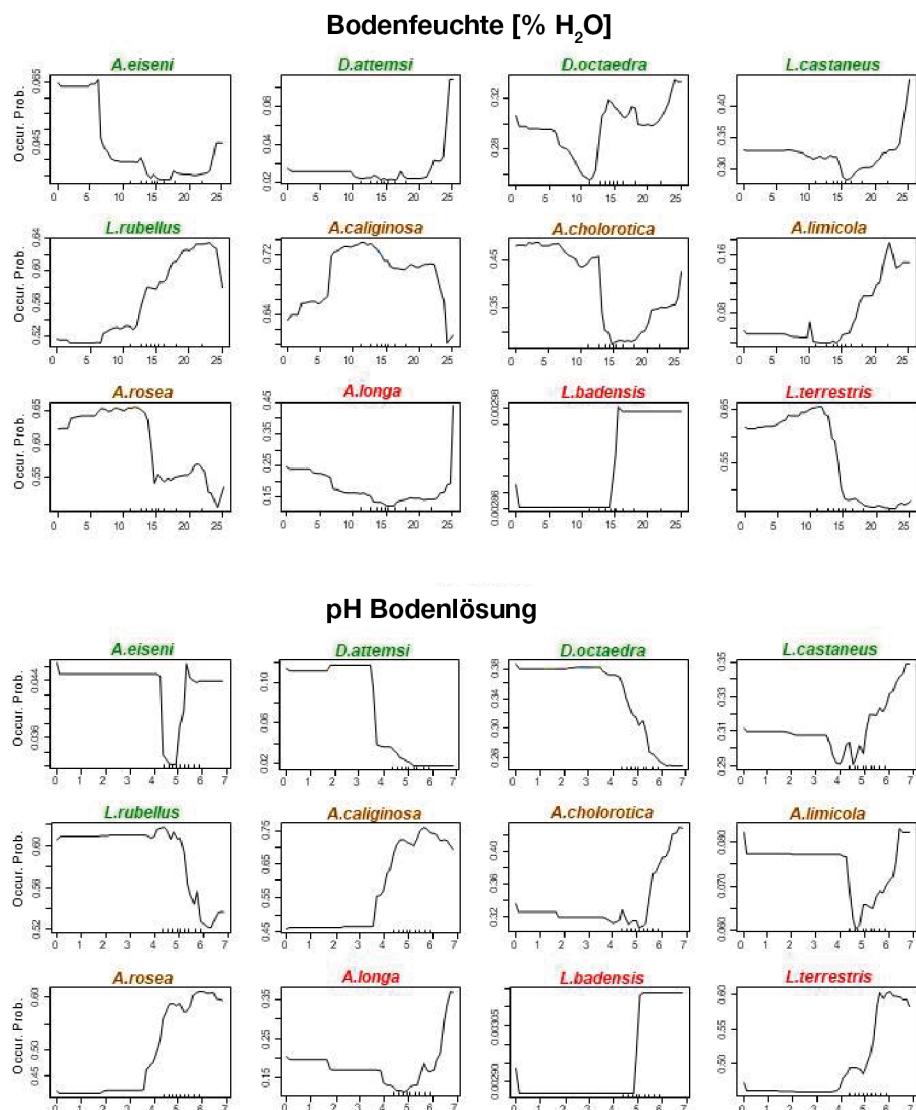
Art	Primäres Habitat	Weitere, sekundäre Habitate
<i>A. caliginosa</i>	Wald, Grünland, Acker	fast alle Habitattypen, außer Meeresinseln und Küsten
<i>A. chlorotica</i>	Acker, Grünland	
<i>A. eiseni</i>	Wald	Grünland, Acker
<i>A. limicola</i>	Wald, Auen, Grünland	Hoch- und Niedermoor
<i>A. longa</i>	Acker, Wald, Grünland	urbane und Bergbaufolgehabitare
<i>A. rosea</i>	Wald, Acker, Grünland	Auen, urbane und Bergbaufolgehabitare
<i>D. attemsi</i>	Acker	Wald
<i>D. octaedra</i>	Wald	schütttere Vegetation
<i>L. badensis</i>	Wald	
<i>L. castaneus</i>	Wald, Auen	Acker, urbane und Bergbaufolgehabitare
<i>L. rubellus</i>	Wald, Grünland	Acker
<i>L. terrestris</i>	Wald, Grünland	

Abbildung 30: Übereinstimmung der prognostizierten Vorkommenswahrscheinlichkeit von *D. octaedra* (rechts) mit Habitattypen in Deutschland (links). Der Vergleich veranschaulicht das potenzielle Vorkommen der Art, das stark mit zusammenhängenden Waldgebieten assoziiert ist.



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 31: Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten auf die Variablen Bodenfeuchte und pH-Wert (= partielle Abhängigkeiten der Umweltprädiktoren von den prognostizierten Vorkommenswahrscheinlichkeitsverteilungen). Grün markierte Artennamen gelten im Allgemeinen als epigäisch, braun als endogäisch und rot als anözisch

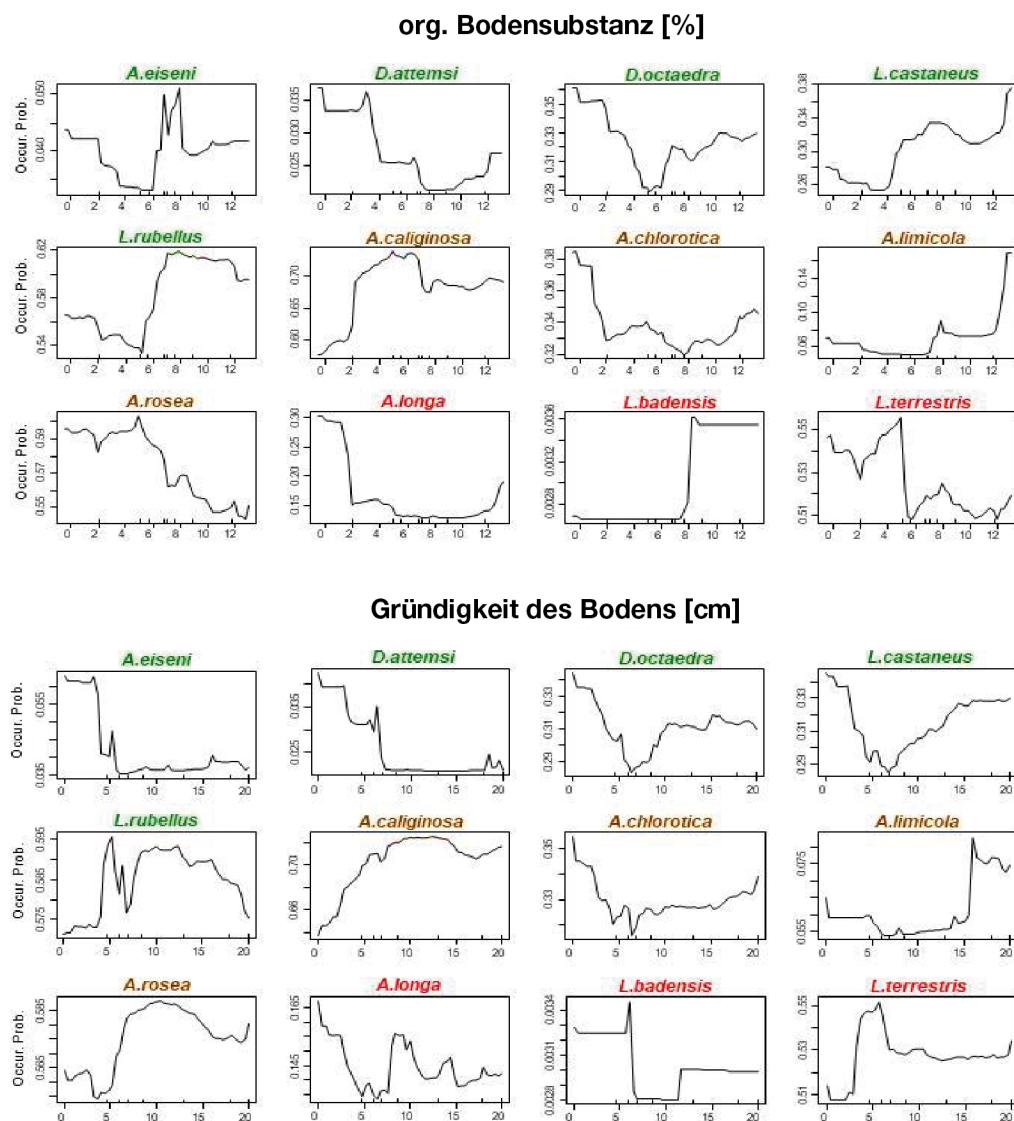


Quelle: Eigene Darstellung, Senckenberg Museum

Die Arten reagieren im Allgemeinen positiv auf einen höheren pH-Wert des Bodens (Abbildung 31), so dass die Wahrscheinlichkeit des Auftretens meist bei höheren pH-Werten zunimmt. Ausnahmen sind die epigäischen Arten *D. attensi*, *D. octaedra* und *L. rubellus*, deren prognostizierter Response bei höheren pH-Werten abrupt abnimmt. Interessanterweise liegt die Schwelle, bei der die Wahrscheinlichkeit des Auftretens entweder zu- oder abnahm, im Allgemeinen bei etwa pH 4,0, mit einem Optimum (für die Arten mit höheren Wahrscheinlichkeiten bei höheren pH-Werten) bei etwa pH 6 – 7. Ausnahmen sind z. B. *A. chlorotica*, *L. rubellus* und *L. terrestris*, deren Schwellenwerte um pH 5 herum liegen, oder *A. longa* mit steigenden Vorkommenswahrscheinlichkeiten oberhalb von Boden-pH-Werten von 6.

Der prognostizierte Response auf den Gehalt an organischer Bodensubstanz (SOM; Abbildung 32) folgt oft keinem einheitlichen Muster und steht nicht im Zusammenhang mit dem Lebensformtyp des Regenwurms. Während die Wahrscheinlichkeit des Vorkommens sowohl einiger epigäischer (d. h. *L. castaneus*, *L. rubellus*) als auch endogäischer Arten (*A. caliginosa*, *A. limicola*) mit höherem SOM-Gehalt zunimmt, nimmt die anderer Arten dagegen ab (epigäisch: *D. attemsi*; endogäisch: *A. chlorotica*, *A. rosea*). Auch anözische Arten zeigen unterschiedlichen Response der Vorhersage bezüglich des Gehalts des Bodens an organischer Substanz, wobei *A. longa* und *L. terrestris* niedrigere und *L. badensis* höhere Vorkommenswahrscheinlichkeiten mit zunehmendem SOM-Gehalt zeigen. Uneinheitlichen Vorhersageresponse zeigen viele andere Arten. Interessanterweise liegen die Schwellenwerte, bei denen sich die Vorhersageantworten ändern – ob negativ oder positiv – im Allgemeinen bei SOM-Niveaus um 6 – 8 %, wobei ein positiver Response oft bei einem Gehalt des Bodens von 12 % SOM ein Maximum erreicht.

Abbildung 32: Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten auf die Variablen organische Bodensubstanz und Gründigkeit des Bodens (= partielle Abhängigkeiten der Umweltprädiktoren von den prognostizierten Vorkommenswahrscheinlichkeitsverteilungen). Grün markierte Artennamen gelten im Allgemeinen als epigäisch, braun als endogäisch und rot als anözisch



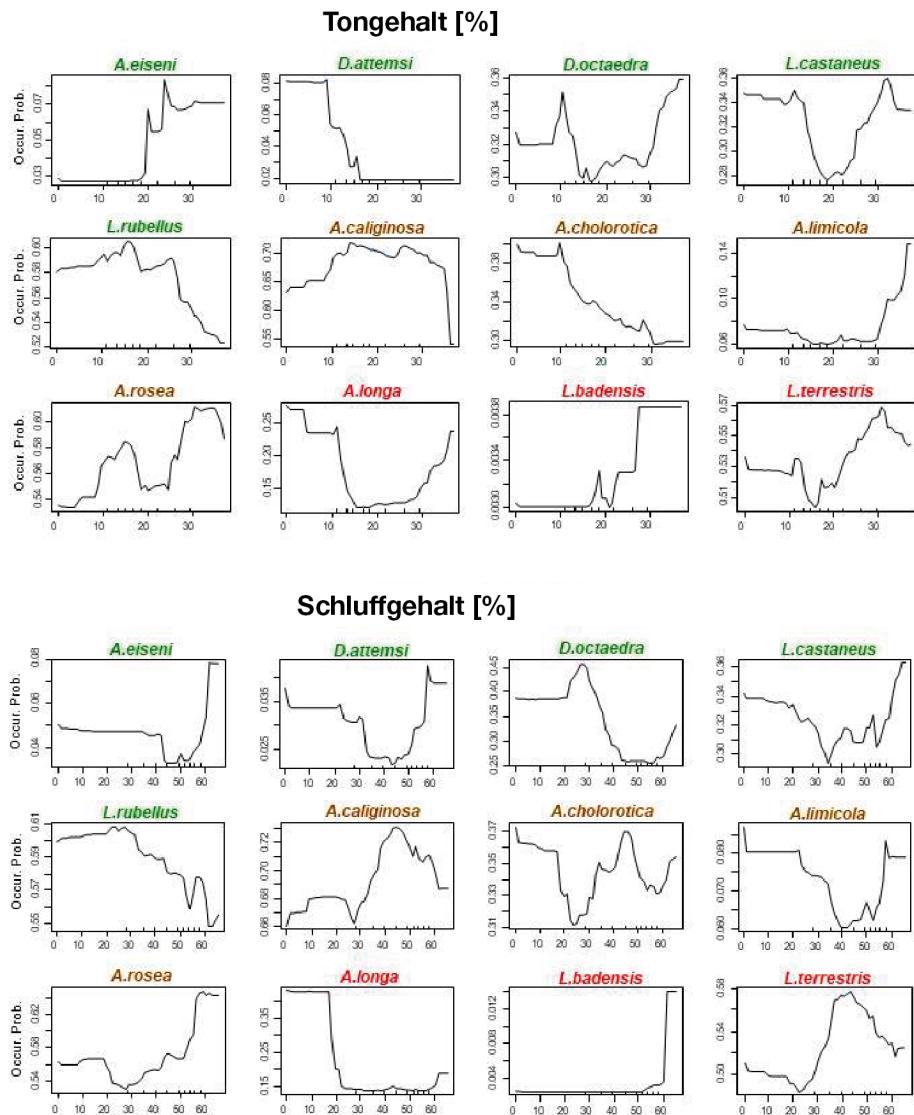
Quelle: Eigene Darstellung, Senckenberg Museum

Da die Gründigkeit des Bodens weniger als 5 % des variablen Beitrags zu den Vorkommensvorhersagen ausmacht, ist der Response der meisten Arten auf diesen Parameter uneinheitlich (Abbildung 32). Dennoch werden für einige epigäische Arten (z. B. *A. eiseni*, *D. attemsi*) negative Vorhersagereaktionen auf zunehmende Gründigkeit beobachtet, während für andere, endogäische Arten (z. B. *A. caliginosa*, *A. rosea*) positive Reaktionen festgestellt wurden. Bei anözischen Arten wurden unregelmäßige Reaktionen auf eine Zunahme der Gründigkeit festgestellt.

Bodentexturparameter (Ton- und Schluffgehalt) tragen nur bei wenigen Arten wesentlich zu den Vorkommensvorhersagen bei (siehe oben). Dementsprechend sind primär nur die Vorhersageantworten dieser Arten auf diese Parameter auffällig (Abbildung 33). Beispielsweise steigen die vorhergesagten Vorkommenswahrscheinlichkeiten mit höheren Ton- und Schluffgehalten für *A. eiseni*, *A. rosea* und *L. terrestris*, was darauf hindeutet, dass diese Arten Böden mit feinerer Textur bevorzugen. Obwohl der Beitrag zu den Vorkommensvorhersagen für *L. badensis* gering war, zeigte diese Art ähnlich positive Vorhersagereaktionen auf Böden mit feiner Textur. Die Ergebnisse für *D. octaedra* sind weniger eindeutig, dennoch ist der Vorhersageresponse bei höherem Tongehalt zunehmend, jedoch bei höherem Schluffgehalt abnehmend, was auch auf eine Bevorzugung feiner texturierter Böden hindeutet. Im Gegensatz dazu nimmt die Vorkommenswahrscheinlichkeit von *A. longa* und *L. rubellus* im Allgemeinen mit höheren Ton- und Schluffgehalten ab, und die von *D. attemsi* nimmt mit höherem Tongehalt ab, mit höherem Schluffgehalt jedoch zu, was auf eine Präferenz für gröbere Böden hindeutet.

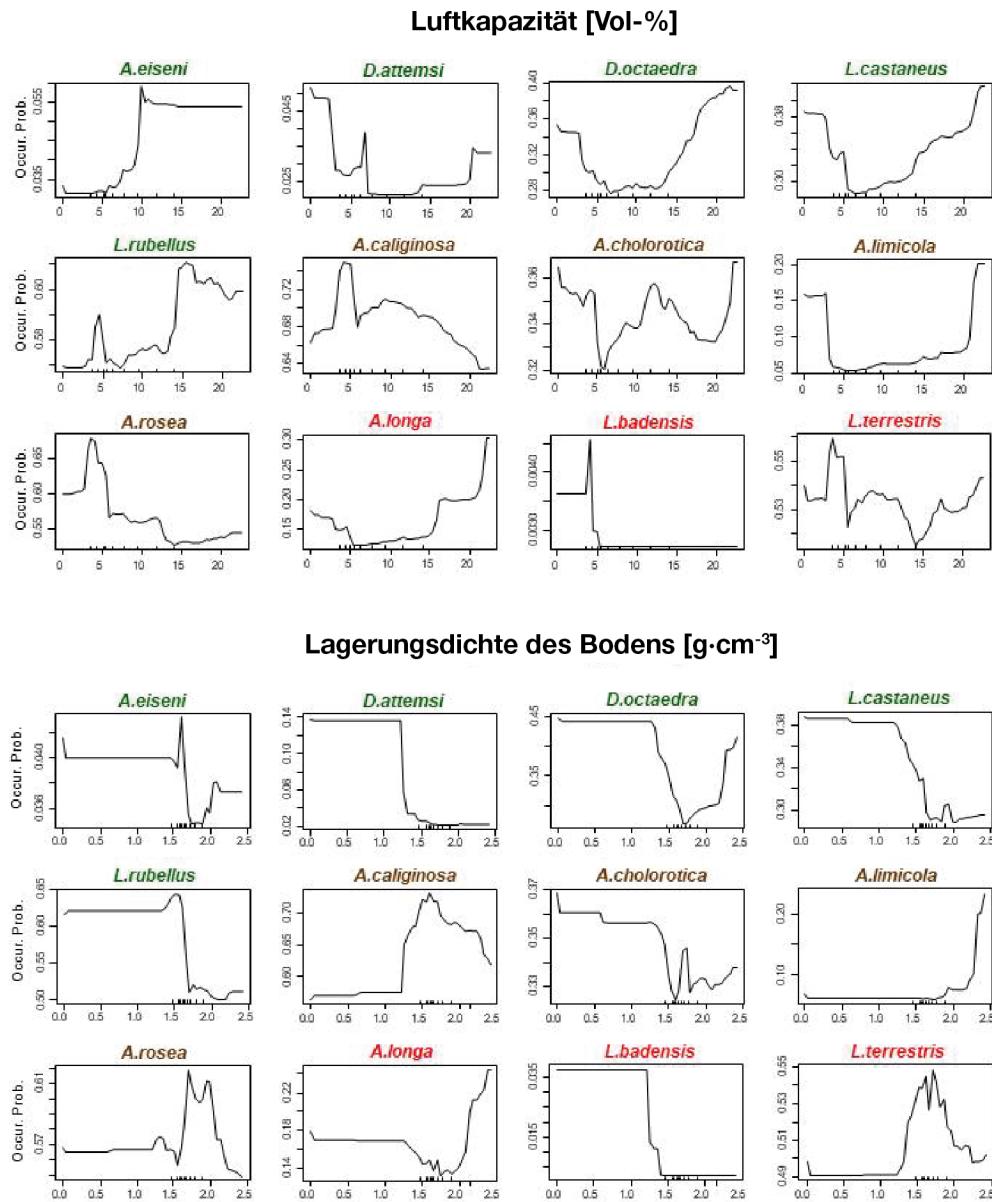
Hinsichtlich der Bodenstruktur zeigten die meisten Arten mit zunehmender Lagerungsdichte (d. h. auch Verdichtung, Abbildung 34) eine negative Reaktion in ihren prognostizierten Vorkommenswahrscheinlichkeiten. Interessant sind *A. limicola* und *A. longa*, die in verdichteten Böden sogar höhere Vorkommenswahrscheinlichkeiten aufweisen. Auch für *A. caliginosa*, *A. rosea* und *L. terrestris* existiert ein Optimum bei recht hoher Lagerungsdichte, dennoch nehmen die prognostizierten Vorkommen bei den am stärksten verdichteten Böden wieder ab. Die Reaktionen auf die Luftkapazität sind im Allgemeinen für alle Arten uneinheitlich. Interessant sind die U-förmigen Reaktionskurven einiger Arten (z. B. *D. attemsi*, *D. octaedra*, *L. castaneus*, *A. limicola*, *A. longa*), die auf geringere Vorkommenswahrscheinlichkeiten bei mittlerer Luftkapazität hindeuten (zwischen ca. 5 und 15 % oder höher).

Abbildung 33: Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten von Arten („Auftrittswahrscheinlichkeit“) auf den Ton- und Schluffgehalt des Bodens als Proxy für die Bodentextur (= partielle Abhängigkeiten der Umweltprädiktoren von den vorhergesagten Vorkommenswahrscheinlichkeitsverteilungen). Grün markierte Artennamen gelten im Allgemeinen als epigäisch, braun als endogäisch und rot als anözisch



Quelle: Eigene Darstellung, Senckenberg Museum

Abbildung 34: Responsekurven der prognostizierten Vorkommenswahrscheinlichkeiten („Vorkommenswahrscheinlichkeit“) von Arten auf die Bodenluftkapazität (Porosität) und die Lagerungsdichte des Bodens als Parameter der Bodenstruktur (= partielle Abhängigkeiten der Umweltprädiktoren von den prognostizierten Vorkommenswahrscheinlichkeitsverteilungen). Grün markierte Arten gelten allgemein als epigäisch, braun als endogäisch und rot als anözisch



Quelle: Eigene Darstellung, Senckenberg Museum

4 Diskussion

4.1 Prognosen

In dieser Studie wird die räumliche Verteilung von Regenwurmgemeinschaften (Gesamtdichte, Artenreichtum) sowie die Wahrscheinlichkeit des Vorkommens und die Größe des geografischen Verbreitungsgebiets von zwölf ausgewählten Arten in Deutschland unter Verwendung mehrerer Modellalgorithmen vorhergesagt, um auf Grundlage der jeweiligen Modellleistung das geeignete Modell für die Projektion der Regenwurmverbreitung in ganz Deutschland auszuwählen. Die Studie folgt den ODMAP-Standardprotokollen für die Implementierung von Artenverbreitungsmodellen (SDMs) (Zurell *et al.*, 2020). Darüber hinaus wurden mehrere Methoden (Experturteil, statistische Verfahren, ökologische Relevanzanalyse) zur Auswahl von Umweltprädiktoren verwendet, um eine Verzerrung bei der Variablenauswahl zu vermeiden und so ein breites Spektrum relevanter Umweltvariablen für eine robuste Modellierung der Verbreitung von Regenwurmarten abzudecken.

Algorithmen des maschinellen Lernens, wie Random Forests (RF), Generalized Boosted Regression Models (GBM) oder die Maximum-Entropie-Methode (MAXENT) haben in jüngster Zeit bessere Ergebnisse erzielt als traditionelle Regressionsmodelle, wie Generalized Linear Regression Models (GLM) oder Generalized Additive Regression Models (GAM) (Elith *et al.*, 2006; Li & Wang 2013; Valavi, 2022). Obwohl RF bisher nur selten verwendet wurden und ihr Potenzial in SDMs nicht ausreichend genutzt wurde, hat ihre hohe Vorhersageleistung in letzter Zeit in angewandten ökologischen Studien Aufmerksamkeit erregt (z. B. Mi *et al.*, 2017). RF und GBM als Ensemble-Klassifikatoren bestehen aus mehreren alternativen Bäumen und verwenden diese bei der Entscheidungsfindung während der Erstellung von Modellvorhersagen (Li & Wang, 2013; Guisan *et al.*, 2017). Obwohl in dieser Studie die Leistung von GLM und GAM übertroffen wurde, sagte GBM nur innerhalb des dritten Quartils der Daten zur Gesamtabundanz und zum Artenreichtum voraus. Der Vergleich der Anpassungsgüte-Statistiken (R^2 , CI, AUC, Kappa), der beobachteten im Vergleich zu den vorhergesagten Datenanpassungen sowie der daraus resultierenden Karten der vorhergesagten Abundanz und des Artenreichtums durch alle Modelle veranschaulicht die gute Leistung von RF. So war dieser Algorithmus beispielsweise in der Lage, Vorhersagen über den Bereich des dritten Quartils der Dichtefelddaten hinaus zu treffen, einschließlich maximaler Dichten von über 600 Ind m^{-2} , und eine große Anzahl von Arten in den Daten zu erfassen. Frühere Studien bestätigen unsere Feststellung, dass Random-Forest-Algorithmen die beste Vorhersageleistung zeigen (z. B. Marmion *et al.*, 2008; Mi *et al.*, 2017; Valavi *et al.*, 2022). Es wurde jedoch eine Verbesserung der Anpassungsgüte für Dichtevorhersagen in RF sichtbar, nachdem zusätzliche Daten (aus Bayern) einbezogen wurden, was bestätigt, dass RF datenabhängig sein kann (s. a. Valavi *et al.*, 2021; Yiu, 2021). Die resultierenden partiellen Responsekurven, die die Beziehungen zwischen Gemeinschaften (oder Arten) und der Umwelt offenlegen, sind ein weiteres Beispiel dafür, wie RF-Modelle ökologisch relevante Ergebnisse liefern (Cutler *et al.*, 2007; Mi *et al.*, 2017).

Trotz der hohen Anpassungsgüte der RF-Modelle ist jedes Vorhersagmodell nur so gut wie die zugrunde liegenden Daten, die zur Kalibrierung verwendet werden. Mit über 20 000 Datensätzen von fast 1000 Fundorten können die biologischen Hintergrunddaten als umfangreich und ausreichend angesehen werden. Nur gut 40 % der Regenwurmdatensätze enthielten auch Daten zu den Umweltprädiktoren, mit einer entsprechend hohen ökologischen Interpretierbarkeit der Funde. Die entsprechend großen Datenlücken mussten durch externe Daten ergänzt werden. Dies war durchaus kritisch für einige Bodenparameter, für die flächenhafte Daten weiträumig auf Interpolationen beruhen (was aufgrund der großen kleinräumigen Heterogenität der Böden ungeeignet ist). Für diese Studie wurden mit großem Aufwand Bodendaten aus kleinräumigen

Erhebungen für diese Studie zusammengetragen. Allerdings konnten nicht alle relevanten Parameter erfasst werden. So halten Creamer *et al.* (2019, in Baritz *et al.*, 2021) Indikatoren für die Qualität der organischen Bodensubstanz (z. B. C/N-, N/P-Verhältnis) auch für Bodenorganismen für sehr wichtig. Entsprechende Daten sind für Deutschland nicht flächendeckend verfügbar.

Es muss auch auf die potenziellen Schwierigkeiten bei der Verwendung externer Habitatdaten hingewiesen werden, da eine zeitliche Trennung zwischen Regenwurmbeobachtungen und Habitattypenübersichten eine auf Landnutzungsänderungen zurückgehende Fehlzuweisung enthalten kann. Glücklicherweise war der Habitattyp die häufigste Umweltmetadatengruppe, die mit den Regenwurmdaten erfasst wurde, so dass eine insgesamt befriedigende Modellkalibrierung gewährleistet war. Schließlich wurden nur abiotische Variablen als Prädiktorvariablen berücksichtigt; etwaige Wechselwirkungen mit anderen Organismen (d. h. zwischen Regenwurmarten, anderer Bodenfauna oder Mikroorganismen) wurden nicht berücksichtigt, da auch für andere Organismen(gruppen) keine entsprechenden Daten verfügbar sind (und Studien über mehrere Bodenorganismengruppen an denselben Standorten äußerst selten sind). Obwohl nicht alle potenziellen Einflussfaktoren auf die Verbreitung von Regenwürmern berücksichtigt werden konnten, enthielten die Modelle eine große Anzahl der wichtigsten Umweltparameter, von denen bekannt ist, dass sie die Fitness von Regenwürmern beeinflussen (z. B. Lee, 1985; Edwards & Arancom, 2022).

Obwohl die Modellvorhersagen noch nicht im Freiland validiert wurden (hier können die im Rahmen von Monitoringprogrammen des Bundes und der Länder vorgesehenen Erhebungen ein erster Schritt sein), bestätigen die veröffentlichten Zusammenstellungen der Autökologie der Regenwurmarten die Mehrzahl der Vorhersagen. Bemerkenswert ist der vorhergesagte Response der Arten auf den pH-Wert des Bodens, wobei viele acidophobe und einige acidophile oder -tolerante Arten mit einem Schwellenwert zwischen pH 4 und 5 ermittelt wurden. Graefe & Beylich (2003) berichten ebenfalls über eine solch starke artspezifische Differenzierung mit einem gemeinsamen Schwellenwert von pH 4,2, mit Ausnahme z. B. der acidophoben *A. longa* mit einem Schwellenwert von pH 5, der auch von unseren Modellen vorhergesagt wurde. Unsere Vorhersagen zu den Reaktionen der Arten auf den Säuregrad des Bodens werden auch weitgehend bestätigt, z. B. von Sims & Gerard (1999), Jänsch *et al.* (2013), Krück (2018) und Sherlock (2018). Diese Autoren sowie Römbke *et al.* (2018) und Lehmitz *et al.* (2016) beschreiben auch artspezifische Präferenzen für den Gehalt an organischer Bodensubstanz (SOM), die durch die Modellvorhersagen fast vollständig bestätigt werden. Einige dieser Autoren berücksichtigten auch Präferenzen für den Tongehalt, die im Allgemeinen, aber nicht immer, von den hier vorgelegten Modellvorhersagen bestätigt werden. So berichten Jänsch *et al.* (2013) über die Vorliebe von *D. octaedra* für Böden mit niedrigem Tongehalt und die leichte Vorliebe von *A. chlorotica* für Tonböden, die beide durch unsere Ergebnisse widerlegt werden. Auch die von uns vorhergesagte positive Reaktion von *L. terrestris* auf Böden mit geringerem Ton- und Schluffgehalt steht im Widerspruch zu der Einschätzung von Sims & Gerard (1999) und Sherlock (2018), dass diese Art lehmreiche Böden bevorzugt (diese Autoren beziehen sich allerdings auf britische Populationen).

Die vorliegende Studie bestätigt die Aussagen früherer Arbeiten zu den Auswirkungen von Niederschlag und Bodenfeuchtigkeit; diese Variablen sind für die Zunahme der Population und die Verteilung erwachsener Regenwürmer nach Regenfällen verantwortlich, insbesondere in relativ trockenen Gebieten oder nach langen Dürreperioden (Lavelle, 1978; Lavelle & Spain, 2005; Kalu *et al.*, 2015; Rajwar *et al.*, 2022). Phillips *et al.* (2019) ermitteln mit einfacheren statistischen Methoden das Klima (insbesondere den mittleren Jahresniederschlag und die Temperatur) als fast ausschließlichen Einflussfaktor auf die Regenwurmgemeinschaften (Gesamtdichte, Artenreichtum) auf globaler Ebene. Unsere Studie bestätigt zwar die kombinierte Rolle von Temperatur,

Niederschlag und Bodenfeuchtigkeit, identifiziert aber auch den Lebensraumtyp, den pH-Wert des Bodens und die organische Bodensubstanz als wichtige Einflussfaktoren, die zusammen etwa 70 % zu den Vorhersagen beitragen. Da das Klima auf globaler Ebene z. B. auch die natürliche Vegetation (die zu Biomen und Lebensraumtypen führt) und teilweise auch die Bodenenge- nese beeinflusst, ist es plausibel, dass statistische Methoden das Klima gegenüber anderen Um- weltparametern als treibende Kraft auf globaler Ebene identifizieren. Unsere Studie auf regionaler Ebene zeigt jedoch auch die Bedeutung von Lebensraum- und Bodenparametern als zusätzliche Faktoren für die Verbreitung von Regenwürmern. Im Gegensatz dazu bezweifeln Hodson *et al.* (2021), dass Bodenfaktoren ein wichtiger Faktor für die Regenwurmdichte und -verteilung sind, insbesondere auf nationaler Ebene. Die maßgeblichen Faktoren für die biologische Vielfalt der Regenwürmer sind also offensichtlich skalenabhängig; Klimaparameter sind auf globaler und regionaler Ebene wichtig, während Lebensraum- und Bodenfaktoren auf kleineren räumlichen Skalen an Bedeutung gewinnen. Auf lokaler Ebene (wo ähnliche klimatische Bedingungen herrschen) dürften die letztgenannten Faktoren am wichtigsten sein, und auch anthropogene Landnutzungsmaßnahmen werden die Artenvielfalt der Regenwürmer zunehmend beeinflussen.

Die vorliegende Studie sagt Vorkommenswahrscheinlichkeiten jenseits der traditionellen Nut- zungstypen Wald, Grünland und Acker vorher, indem alle terrestrischen EUNIS-Level-1-Lebens- raumtypen, einschließlich mariner (d. h. Inseln), küstennaher und städtischer Standorte, einbe- zogen wurden. Diese Vorhersagen stimmten in hohem Maße mit den vorgeschlagenen Klassifi- zierungen der Verbreitungsgebiete auf der Grundlage der Verbreitungsvorhersagen überein. So wurde für die meisten Arten mit großem Verbreitungsgebiet eine weite Verbreitung in vielen Regionen Deutschlands und ein Vorkommen in vielen verschiedenen Lebensraumtypen pro- gnostiziert, oft mit Wahrscheinlichkeiten von mehr als 50 – 60 %, was auf ihren ökologisch gene- ralistischen Charakter hinweist. In der Roten Liste Deutschlands werden alle diese Arten als sehr häufig aufgeführt (Lehmitz *et al.*, 2016), und in der Literatur werden sie als euryök bzw. wird ihr Vorkommen als ubiquitär beschrieben (z. B. Sims & Gerard, 1999; Jänsch *et al.*, 2013; Römbke *et al.*, 2018; Sherlock, 2018), was unsere Ergebnisse bestätigt. Obwohl *L. terrestris* im Allgemeinen als eurytop angesehen wird, wird gelegentlich auf eine leichte Präferenz für Grün- landstandorte hingewiesen (Sims & Gerard, 1999; Jänsch *et al.*, 2013; Sherlock, 2018), was durch unsere Vorhersagen bestätigt wird, die aber in sehr ähnlichem Maße auch Waldhabitatem vorher- sagen. Die Art gilt als störungsintolerant (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018), was ihre ge- ringe Wahrscheinlichkeit für natürliche Sonderstandorte (z. B. Auen, Moore) sowie anthropogen stark beeinflusste Lebensraumtypen erklären könnte. Vorsicht ist u. U. bei den Aussagen zu *A. caliginosa* geboten, die kryptische Arten enthalten kann (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018); dies ist jedoch eine Frage der Taxonomie und der Artbestimmung und nicht der Modellie- rungsverfahren.

Den von uns als Arten mit mittlerem Verbreitungsgebiet eingestuften Arten wird ebenfalls ein weites Vorkommen in Deutschland vorhergesagt, wenn auch oft mit deutlich geringerer Wahr- scheinlichkeit. In der Roten Liste Deutschlands werden sie alle als häufig geführt. Obwohl für sie ein Vorkommen in vielen verschiedenen Lebensraumtypen prognostiziert wird (euryöke Arten), scheinen sie stärker habitatabhängig zu sein und ein Optimum in bestimmten Lebensraumtypen zu haben. So wurde beispielsweise für *A. chlorotica* ein stärkeres Vorkommen auf landwirt- schaftlich genutzten Flächen (Acker oder Grünland) vorhergesagt, was auch aus Beobachtungs- daten hervorgeht (z. B. Jänsch *et al.*, 2013; Römbke *et al.*, 2018). Andererseits wird *D. octaedra* hauptsächlich in Waldhabitaten vorkommend vorhergesagt, was mit Erkenntnissen von z. B. Jänsch *et al.* (2013), Römbke *et al.* (2018) und Sherlock (2018) korrespondiert. In Anbetracht der acidophilen Natur der Art ist eine Präferenz für Nadelwälder wahrscheinlich (vgl. Sherlock, 2018). Während *A. castaneus* eher ein Generalist zu sein scheint, zeigen die vorliegenden Ergeb- nisse die höchsten Vorkommenswahrscheinlichkeiten in Wäldern und Auen, was u. a. Jänsch *et*

al. (2013), Römbke *et al.* (2018) und Krück (2018) widerspricht, die eine Präferenz auch für Grasland sehen. Interessanterweise wurde *A. longa* von unseren Modellen auch als Generalist vorhergesagt, der in verschiedenen Lebensraumtypen vorkommt, aber in feuchteren Lebensräumen (z. B. Inseln, Küsten, Überschwemmungsgebieten, Mooren) fehlt. Dies wird von Krück (2018) bestätigt, der *A. longa* eine Vorliebe für trockenere Lebensräume bescheinigt, im Gegensatz zu Sims & Gerard (1999), die ein Vorkommen der Art in Überschwemmungsgebieten Großbritanniens feststellten.

Die Arten mit eingeschränktem Verbreitungsgebiet weisen alle eine auf bestimmte Regionen und Lebensraumtypen beschränkte Vorkommenswahrscheinlichkeit auf. In der Roten Liste Deutschlands sind sie alle als *selten* oder *sehr selten* aufgeführt. Zum Beispiel liegen die höchsten Verbreitungswahrscheinlichkeiten von *D. attemsi* eher in den Hügelländern und Mittelgebirgen; die höchsten Wahrscheinlichkeiten wurden für die Nutzungsart Acker (und in zweiter Linie für Wälder) vorhergesagt. Dies steht im Widerspruch zu früheren Arbeiten, mit Ausnahme von Sherlock (2018), der Vorkommen von *D. attemsi* in Wäldern Großbritanniens feststellte. *A. eiseni* wird von den Modellen als am wahrscheinlichsten in Wäldern (in Mittel- und Süddeutschland) vorkommend vorhergesagt, wie auch Römbke *et al.* (2018) und Lehmitz *et al.* (2016) berichten. *A. limicola* ist als hydrophil bekannt (Sims & Gerard 1999; Lehmitz *et al.* 2016; Krück 2018; Römbke *et al.*, 2018; Sherlock, 2018). Dementsprechend sagen die Modelle voraus, dass die Art mit hoher Wahrscheinlichkeit in Überschwemmungsgebieten sowie im Grasland und in Wäldern vorkommt – die sich möglicherweise ebenfalls in Überschwemmungsgebieten u. ä. befinden, deren Funddaten aber möglicherweise fälschlich allgemeineren Lebensraumtypen zugeordnet wurden. Die Modelle sagen voraus, dass *A. limicola* vor allem in Westdeutschland (und am stärksten im Rheintal) vorkommt, was zu den Befunden von Krück (2018) passt, dass die Art in Nordostdeutschland selten vorkommt. *L. badensis* ist eine in Deutschland endemische Art, die in den Wäldern des Hochschwarzwalds (Südwestdeutschland) vorkommt (Lehmitz *et al.*, 2016), was sich auch in den Modellvorhersagen zeigt.

Interessant sind die wenigen Arten, die nach den Modellierungen in Ausnahmehabitaten vorkommen. So wurde für *A. limicola*, *L. rubellus* und *L. castaneus* ein Vorkommen in marin beeinflussten Lebensräumen (d. h. auf Inseln) und für *A. caliginosa*, *A. chlorotica* und *A. limicola* in Küstengebieten vorhergesagt; alle jedoch mit einer geringen (<35 %) Vorkommenswahrscheinlichkeit. Das weist auf ein lückenhaftes Vorkommen in diesen Lebensräumen hin. Auffallend ist die große Anzahl von Arten, die mit mäßiger Wahrscheinlichkeit in städtischen, industriellen und anderen anthropogenen Gebieten vorkommen. Diese neuen Ergebnisse können bei der Bewertung von Erhebungen zur biologischen Vielfalt des Bodens in solchen Gebieten helfen.

Auf der Ebene der Lebensgemeinschaften ist die geografische Variation und die inverse Beziehung zwischen Artenreichtum und Gesamtabundanz, wie sie in dieser Studie (hauptsächlich für Nordostdeutschland) festzustellen ist, ein häufiges Phänomen in der Ökologie. Gesamtabundanz und Artenreichtum verhalten sich oft gegenläufig, so dass ein Gebiet eine hohe Individuendichte, aber einen geringen Artenreichtum aufweist (Verberk *et al.*, 2011). Eine Region mit ungünstigeren Umweltbedingungen ermöglicht ggf. nur das Vorkommen einiger weniger Arten, diese aber in durchaus großen Populationen (hohe Individuenzahlen), u. U. gerade aufgrund geringerer Konkurrenz durch andere Arten (Groves, 2022). Dies könnte möglicherweise die hohe Individuendichte, aber den geringen Artenreichtum erklären, der für Nordostdeutschland (insbesondere Brandenburg) vorhergesagt wird, wo vergleichsweise trockenere, sandige Böden vorherrschen und wo vorwiegend Nadelbaumforste stocken. In diesem Zusammenhang sind die hohen Vorkommenswahrscheinlichkeiten von *D. octaedra* und *L. rubellus* in diesen Gebieten auffällig, bei des epigäische, acidophile (oder -tolerante) Arten mit einer vermuteten Affinität zu Wäldern. Eigene Erhebungen der biologischen Vielfalt im Boden haben oft hohe Populationsdichten von

sehr wenigen epigäischen Arten in Wäldern auf sandigen Böden gezeigt. Andererseits werden für die bayerischen Alpen und das Rheintal als einige der wenigen Regionen mit hoher Regenwurm-Biodiversität (sowohl Gesamtabundanz als auch Artenreichtum) in Deutschland vorhergesagt. Das Rheintal ist für reiche Böden und eine hohe allgemeine Biodiversität bekannt, und die Vorhersagen für die alpinen Regionen gehen konform mit der Alpenkonvention (2014), die die Alpen aufgrund ihres natürlichen Zustands und der geringe(re)n menschlichen Störungen als eine der reichsten Regionen Europas in Bezug auf die Vielfalt von Pflanzen und Tieren anerkennt.

Die Größe des Verbreitungsgebiets ist seit langem als guter Indikator für die Bewertung der Bedrohung einer Art anerkannt. Der Bedrohungsstatus der Regenwürmer in Deutschland ist in der entsprechenden Roten Liste dokumentiert (Lehmitz *et al.*, 2016). Die Kartierung der räumlichen Verbreitung der Regenwurmarten und die Bestimmung der Flächengröße erlauben, die Arten in Verbreitungsgruppen (Arten mit großem, mittlerem und kleinem Verbreitungsgebiet) einzuteilen und Aussagen zum Bedrohungsstatus und Schutzbedarf der Arten abzuleiten (IUCN 2012 a, b; 2022).

Keine Art wurde nach den Kriterien des Verbreitungsgebiets als bedroht eingestuft (die Verbreitungsgebiete aller Arten überstiegen den Mindestschwellenwert von 20 000 km²), was darauf hindeutet, dass die untersuchten Regenwurmarten in Deutschland potenziell weit genug verbreitet sind, als dass sie durch Einzelereignisse aussterben (IUCN 2012, a, b). Dennoch zeigt ein Vergleich der prognostizierten Verbreitungskarten und des berechneten Verbreitungsgebiets (Area of Occupancy, AOO), dass bestimmte Arten aufgrund ihres eingeschränkten Verbreitungsgebiets zur Anlass zur Sorge geben, z. B. *L. castaneus*, *D. octaedra* und *A. longa* (als gefährdete Arten) sowie *A. eiseni*, *D. attemsi* und *A. limicola* (als stark gefährdete Arten nach den AOO-Kriterien) oder weil sie in Deutschland endemisch sind, wie *L. badensis*. Über den Status von *A. eiseni* in Deutschland existieren unterschiedliche Meinungen; während Bouche (1972) und Graff (1953) die Art in Frankreich bzw. Deutschland als selten einstufen, stufen Römbke *et al.* (2017) die die Art als häufig ein. Unsere Ergebnisse unterstützen teilweise die älteren Befunde von Graff (1953) und Bouche (1972) sowie die Zwischenposition von Lehmitz *et al.* (2016), die die Art als mäßig häufig einstuften (die vorliegende Studie bevorzugt den Begriff des eingeschränkten Verbreitungsgebiets).

Dagegen müssen die hier gemachten Vorhersagen über ein begrenztes Vorkommen von *A. eiseni* in hessischen Wäldern (Römbke *et al.*, 2017) und einigen anderen Gebieten zurückhaltend beurteilt werden. Die Art wird wahrscheinlich arboreal und kortikal, und die begrenzten Beobachtungsdaten (auf denen die Modelle basieren) könnten methodisch verzerrt sein, da die üblichen Methoden zur Regenwurmextraktion die bevorzugten Mikrohabitatem dieser Art möglicherweise nicht ausreichend erfassen (Lehmitz *et al.*, 2016; Römbke *et al.*, 2018). *A. limicola* ist die einzige hier untersuchte Art, die in der deutschen Roten Liste der Regenwürmer als gefährdet aufgeführt ist. Sein prognostiziertes Vorkommen in Feuchtgebieten, das seinen Status als hydrophile Art bestätigt, sowie seine prognostizierte begrenzte Verbreitung in Deutschland spiegeln seinen Gefährdungsstatus wider. Die übrigen Arten sollten bei künftigen Erhebungen zur biologischen Vielfalt im Boden als Schwerpunktarten betrachtet werden.

Besonders hervorzuheben ist *L. badensis*, eine endemische Art, die in der Hochschwarzwaldregion vorkommt und wahrscheinlich gefährdet ist (Lehmitz *et al.*, 2016). Obwohl keine ausreichenden Funddaten für die Berechnung von EOO oder AOO zur Verfügung standen, wurde für diese Art ein sehr eingeschränktes und enges Verbreitungsgebiet vorhergesagt, was ihrer Einstufung als gefährdet untermauert. Diese Art verdeutlicht auch einen wichtigen Aspekt der Verbreitungsmodellierung: Obwohl die Modelle ein potenzielles Vorkommen z. B. in den bayeri-

schen Alpen vorhersagen, wurde die Art dort noch nie nachgewiesen. Die Modellergebnisse zeigen demnach nur eine hohe potenzielle Lebensraumeignung für die Art in den Alpen, sind aber kein Beleg für ihr dortiges Vorkommen.

4.2 Daten

Eine Erweiterung auf andere Bodentiergruppen als die Lumbriciden erscheint aufgrund fehlender Funddaten aus großen Regionen Deutschlands nicht möglich.

Da die Auswertung von Bodentierfunddaten nur standortsbezogen und damit am Punkt möglich. Die Verknüpfung der Bodentierfundpunkte mit Standortinformation geschieht am besten mit am Fundpunkt. Eine Ergänzung fehlender Standortinformation aus Bodenflächendaten ist insofern grundsätzlich problematisch, weil letztere flächenrepräsentativ für die jeweilige Legendeneinheit sind, also am Punkt nur begrenzte Aussagekraft haben. Die Flächendaten sind für eine punktuelle Abfrage unter Umständen zu stark generalisiert. Dies führt zu Unschärfen bei der statistischen Berechnung und der Modellierung von Verbreitungskarten. Hinzu kommt, dass in den Flächen-Bodenprofilen häufig klassierte statt diskreter Werte verwendet werden. Die Kombination von nominal- und ordinalskalierten Daten mit kardinalskalierten Daten kann problematisch sein.

Schlussfolgerungen

Die vorliegende Studie ist die erste umfassende Analyse, die die Verteilung von Regenwürmern auf deutschlandweiter Ebene modelliert und hierfür die wichtigsten Arten und mehrere Umweltfaktoren einbezieht. Frühere Regenwurm-SDM-Studien, die Biodiversitätsmaße, wie Gesamtdichte, Biomasse und Artenreichtum verwenden, verwenden klassische oder traditionelle Regressionsmodelle (GLM, GLMM und GAM). Angesichts des Potenzials der maschinellen Lernmodelle GBM und RF mit hoher Vorhersageleistung wurde in dieser Studie deren Vorhersageleistung mit der traditionellen Regressionsmodelle verglichen, um das beste statistische Modell für die Vorhersage der Regenwurm-Biodiversität in Deutschland zu ermitteln. Die Vorhersageleistung von RF war besonders hoch.

Diese Vorhersagen, einschließlich der Einteilung der Arten nach der Größe ihres Verbreitungsgebiets sowie des gemeinschafts- und artspezifischen Responses auf ein breites Spektrum von Umweltvariablen, bieten eine effektive Annäherung an die Regenwurmverteilung und deren Einflussfaktoren in Deutschland. Diese Informationen sind sowohl wertvoll für zukünftige wissenschaftliche Feldstudien, als auch eine Voraussetzung für Boden-Biodiversitäts-Monitoringprogramme, die standardisierte Vergleichswerte für die Ergebnisbewertung benötigen. Aus den Modellergebnissen sollen Referenzwerte extrahiert werden, die auf spezifischen Standortbedingungen basieren und für den Einsatz in Boden-Biodiversitäts-Monitoringprogrammen vorgesehen sind; ein Werkzeug hierzu ist in Entwicklung.

Entsprechende Monitoringprogramme werden zur Validierung der Modellergebnisse beitragen. Wichtig ist, dass bei den Erhebungen zur Artenvielfalt die Umweltparameter (insbesondere des Bodens) mit erhoben werden, um die Daten zu den Art-Umweltfaktoren und damit die Modellgenauigkeit zu verbessern.

Die Modellierung und flächenhafte Darstellung der Regenwurmverbreitung ermöglicht außerdem, die Arten nach der Größe ihres geografischen Verbreitungsgebiets einzustufen. Das liefert wichtige Grundlagen für Entscheidungen für die Arterhaltung. Die Modellierung der Verbreitung der Arten wurde mithilfe der zehn EUNIS-Lebensraumtypen der Ebene 1 als Prognosekriterium vorgenommen. Dies ist eine deutliche Verbesserung gegenüber der klassischen Einteilung nach den drei

Hauptnutzungen Acker, Grünland und Wald. Die Verbreitung von Regenwürmern innerhalb der Lebensraumtypen der Ebene 1 kann trotz der besseren Differenzierung sehr unterschiedlich sein. So kann beispielsweise Wald unter anderem in Laub-, Nadel- und Mischwälder unterteilt werden. Bewirtschaftungsmaßnahmen in landwirtschaftlich genutzten Lebensräumen haben einen starken Einfluss auf Regenwurmgemeinschaften. Die Verfügbarkeit von hochauflösenden Rasterdaten auf EUNIS-Ebene 2 oder feineren Hierarchien würde die Modellgenauigkeit bezüglich der Lebensraumpräferenzen der Arten erhöhen. Wichtig ist, dass die Erfassung von Lebensraumtypen auf einer differenzierteren Ebene während der Erhebungen zur biologischen Vielfalt des Bodens künftige Datensynthesen und die Kenntnis spezifischer Einflussfaktoren erheblich verbessern wird, so dass bessere Erhaltungsentscheidungen getroffen werden können. Obwohl das Random-Forest-Modell in der Lage war, die potenzielle Verbreitung einiger Arten in weitgehend unbeprobenen Gebieten Norddeutschlands (insbesondere in Mecklenburg-Vorpommern, Niedersachsen und den nördlichen Teilen Nordrhein-Westfalens) vorherzusagen, ohne dass Überanpassungen des Modells an die Trainingsdaten auftraten, kann es Feldbeobachtungen nicht ersetzen. Erhebungen in den bisher nicht beprobenen Gebieten wären außerordentlich hilfreich.

Besonderes Augenmerk sollte auf Arten mit begrenztem Verbreitungsgebiet, wie *D. attemsi*, *A. eiseni* und *L. badensis*, gelegt werden. Bei Arten mit klar definierten Habitatansprüchen, wie *A. limicola* in Feuchtgebieten und *D. octaedra* in Wäldern, sollten die Lebensräume, in denen sie vorkommen, auf mögliche Lebensraumverschlechterungen überwacht werden. Detaillierte Studien über den endemischen *L. badensis* wären überdies hilfreich, um genauere SDMs und Berechnungen der geografischen Verbreitung zu ermöglichen und eine bessere Einschätzung der tatsächlichen Verbreitung und des Schutzbedarfs zu erhalten.

5 Quellenverzeichnis

- Adhikari, K., Hartemink, A. E. (2015): Linking soils to ecosystem services — A global review. In: *Geoderma*, 262, 101–111.
- Aiello-Lammens, M. A., Boria, R. A., Radosavljevic, A., Vilela, B., Anderson, R. P. (2015): spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. In: *Ecography*, 38, 541–545.
- Allouche, O., Tsoar, A., Kadmon, R. (2006): Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS). In: *Journal of Applied Ecology*, 43, 1223–1232.
- Alpine Convention (2014): Guidelines for climate change adaptation at the local level in the Alps. Alpine Signals 7. Permanent Secretariat of the Alpine Convention, Innsbruck, ISBN: 9788897500247, 45 S.
- Amossé, J., Turberg, P., Kohler-Milleret, R.-, Gobat, J.-M., Le Bayon, R.-C. (2015): Effects of endogeic earthworms on the soil organic matter dynamics and the soil structure in urban and alluvial soil materials. In: *Geoderma*, 50–57, 243–244.
- Anderson, J. M. (1975): The enigma of soil animal species diversity. In: VANEK J (ed) *Progress in Soil Zoology*. Dr. W. Junk B. V. Publishers, The Hague, S. 51–58.
- André, H. M., Noti, M.-I., Lebrun, P. (1994): The soil fauna: the other last biotic frontier. In: *Biodiversity and Conservation*, 3:45–56.
- Austin M. P. (2002): Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. In: *Ecological Modelling*, 157, 101–18.
- Bakker, M. R., Brunner, I., Ashwood, F., Bjarnadottir, B., Bolger, T., Børja, I., Carnol, M., Cudlin, P., Dalsgaard, L., Erktan, A., Godbold, D., Kraigher, H., Meier, I. C., Merino-Martín, L., Motiejunaité, J., Mrak, T., Oddsdóttir, E. S., Ostonen, I., Pennanen, T. L., Püttsepp, Ü., Suz, L. M., Vanguelova, E. I., Vesterdal, L., Soudzilovskaia, N. A. (2019): Belowground Biodiversity Relates Positively to Ecosystem Services of European Forests. In: *Front. For. Glob. Change*, 2:6. (<https://doi.org/10.3389/ffgc.2019.00006>).
- Ballabio, C., Panagos, P., Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. In: *Geoderma*, 261, 110–123. (<https://doi.org/10.1016/j.geoderma.2015.07.006>).
- Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. In: *Geoderma*, 355, 113912. (<https://doi.org/10.1016/j.geoderma.2019.113912>).
- Bardgett, R. D. (2005): *The Biology of Soil: A Community and Ecosystem Approach*. Oxford Univ. Press, Oxford, 254 S.
- Baritz, R., Amelung, W., Antoni, V., Boardman, J., Horn, R., Prokop, G., Römbke, J., Romkens, P., Steinhoff-Knopp, B., Swartjes, F., Trombetti, M., de Vries, W. (2021): Soil monitoring in Europe. Indicators and thresholds for soil quality assessments. EEA ETC/ULS Report. European Environmental Agency. (https://www.eea.europa.eu/publications#c7=en&c11=25&c14=&c12=&b_start=0&c13=soil), 137 S.
- Behan-Pelletier, V. M., Bisset, B. (1992): Biodiversity of nearctic soil arthropods. In: *Can. Biodiv.*, 2(3), 5–14.
- BGR [Bundesanstalt für Geowissenschaften und Rohstoffe, in cooperation with the State Geological Surveys of Germany] (1997 – 2018): Bodenübersichtskarte 1 : 200 000 der Bundesrepublik Deutschland. Hannover, Berlin.
- Bhatt, S., Cameron, E., Flaxman, S. R., Weiss, D. J., Smith, D. L., Gething, P. W. (2017): Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of the Royal Society Interface* 14(134): 20170520 (<https://doi.org/10.1098/rsif.2017.0520>).
- Biber, M. F., Voskamp, A., Niamir, A., Hickler, T., Hof, C. (2020): A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. In: *J. Biogeogr.*, 47:114–129.

- Blouin, M., Hodson, M. E., Delgado, E. A., Baker, E. A., Brussaard, L. K., Butt, R., Dai, J., Dendooven, L., Peres, G., Tondoh, J. E., Cluzeau, D., Brun, J.-J. (2015): A review of earthworm impact on soil function and ecosystem services. In: *European Journal of Soil Science*, 64, 161–182.
- Bobrowski, M., Weidinger, J., Schwab, N., Schickhoff, U. (2021): Searching for ecology in species distribution models in the Himalayas. In: *Ecological Modelling*, 458, 109693 (<https://doi.org/10.1016/j.ecolmodel.2021.109693>).
- Boria, A.R, Olson, L. E., Goodman, S. M., Anderson, R. P. (2014): Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. In: *Ecological Modelling*, 275, 73–77.
- Bouché, M. B. (1972): *Lombriciens de France. Écologie etsystématique*. In: *Annales de zoologie. Ecologie animale, hors série*, 72(2): 1–671.
- Boumal, J., Montanarella, L. (2016): Facing policy challenges with inter- and transdisciplinary soil research focused on the UN Sustainable Development Goals. In: *Soil*, 2, 135–145.
- Brambilla, M., Bergero, V., Bassi, E., Falco, R. (2014): Current and future effectiveness of Natura 2000 network in the central Alps for the conservation of mountain forest owl species in a warming climate. In: *European Journal of Wildlife Research*, 61, 35–44.
- Breiman, L. (2001): Random forests. In: *Machine Learning*, 45, 5–32.
- Briones, M. J. I. (2018): The Serendipitous Value of Soil Fauna in Ecosystem Functioning: The Unexplained Explained. In: *Front. Environ. Sci.*, 6, 149, doi: 10.3389/fenvs.2018.00149.
- Brussaard, L. (1998): Soil fauna, guilds, functional groups and ecosystem processes. In: *Applied Soil Ecology*, 9, 123–135.
- Brussaard, L., de Ruiter, P. C., Brown, G. G. (2007): Soil biodiversity for agricultural sustainability. In: *Agriculture, Ecosystems and Environment*, 121, 233–244.
- Capinha, C., Essl, F., Seebens, H., Pereira, H. M., Kühn, I. (2018): Models of alien species richness show moderate predictive accuracy and poor transferability. In: *NeoBiota*, 38, 77–96.
- Cluzeau, D., Guernion, M., Chaussod, R., Martin-Laurent, F., Villenave, C., Cortet, J., Ruiz-Camacho, N., Pernin, C., Mateille, T., Philippot, L., Bellido, A., Rougé, L., Arrouays, D., Bispo, A., Pérès, G. (2012): Integration of biodiversity in soil quality monitoring: Baselines for microbial and soil fauna parameters for different land-use types. In: *European Journal of Soil Biology*, 49, 63–72.
- Cobos, M. E., Barve, V., Barve, N., Jimenez-Valverde, A., Nuñez-Penichet, C. (2021): rangemap: Simple Tools for Defining Species Ranges. <https://cran.r-project.org/web/packages/rangemap/index.html>.
- Crittenden, S. J., Eswaramurthy, T., de Goede, R. G. M., Brussaard, L., Pulleman, M. M. (2014): Effect of tillage on earthworms over short- and medium-term in conventional and organic farming. In: *Applied Soil Ecology*, 83, 140–148.
- Cutler, D. R., Edwards, T. C. Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. (2007): Random Forest classification in ecology. In: *Ecology*, 88, 2783–2792.
- Delgado-Baquerizo, M., Reich, P. B., Trivedi, C. et al. (2020): Multiple elements of soil biodiversity drive ecosystem functions across biomes. In: *Nature Ecology and Evolution*, 4, 210–220 (<https://doi.org/10.1038/s41559-019-1084-y>).
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquèz, J. R. G., Gruber, B., Lafourcade, B., Leitao, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schroder, B., Skidmore, A. K., Zurell, D., Lautenbach, S. (2013): Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. In: *Ecography*, 36(1), 27–46.

- Dunbar, M. B., Panagos, P., Montanarella, L. (2013): European Perspective of Ecosystem Services and Related Policies. In: *Integr. Environ. Assess. Manag.*, 9, 231–236.
- EASAC (2018): Opportunities for soil sustainability in Europe. EASAC policy Report 36. 48 S.
- Edwards, C. A., Arancon, N. Q. (2022): *Biology and Ecology of Earthworms*. 4th Ed., Springer, New York, 567 S.
- Elith, J., Graham, C. H., Anderson, R. P., et al (2006): Novel methods improve prediction of species' distributions from occurrence data. In: *Ecography*, 29, 129–151.
- Elith, J., Leathwick, J. R., Hastie, T. (2008): A working guide to boosted regression trees. In: *Journal of Animal Ecology*, 77, 802–813.
- Faber, J. H., Creamer, R. E., Mulder, C., Römbke, J., Rutgers, M., Sousa, J. P., Stone, D., Griffiths, B. S. (2013): The practicalities and pitfalls of establishing a policy-relevant and cost-effective soil biological monitoring scheme. In: *Integrated Environmental Assessment and Management*, 9(2): 276–284.
- FAO, ITPS, GSBI, CBD, and EC (2020): State of knowledge of soil biodiversity - Status, challenges and potentialities. Report 2020. FAO, Rome (<https://doi.org/10.4060/cb1928en>). 585 S.
- Ferrier, S., Manion, G., Elith, J., Richardson, K. (2007): Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. In: *Diversity Distrib.*, 13, 252–264.
- Fortin, M. J., Dale, M. (2005): *Spatial Analysis A Guide for Ecologists*. Cambridge University Press, Cambridge New York Melbourne, 380 S.
- Fourcade, Y., Engler, J. O., Rodder, D., Seondi, J. (2014): Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. In: *PLoS ONE*, 9(5): e97122. (<https://doi.org/10.1371/journal.pone.0097122>).
- Fox, J. (2015): *Applied Regression Analysis and Generalised Linear Models*. Sage Publications, Thousand Oaks London, 816 S.
- Frund, H.-C., Graefe, U., Tischer, S. (2011): Earthworms as Bioindicators of Soil Quality (Ch. 11). In: Karaca, A. [Hrsg.]: *Biology of Earthworms, Soil Biology* 24. Springer-Verlag, Berlin, Heidelberg, S. 261–278.
- Gardi, C., Jeffery, S. (2009): Soil biodiversity. JRC Scientific and Technical Report 50304. DOI 10.2788/7831.
- Gaston, K. J., Fuller, R. A. (2009): The sizes of species' geographic range. In: *Journal of Applied Ecology*, 46, 1–9.
- GDAL-Software-Suite (2013): Geospatial data abstraction library. <http://www.gdal.org>
- Giller, P. S. (1996): The diversity of soil communities, the “poor man's tropical forest”. In: *Biodiversity and Conservation*, 5:135–168.
- Glæsner, N., Helming, K., de Vries, W. (2014): Do current European policies prevent soil threats and support soil functions? In: *Sustainability*, 6, 9538–9563.
- Global Invasive Species Database (2022): Species profile: *Dendrobaena octaedra*. Downloaded from <http://www.iucngisd.org/gisd/species.php?sc=1710>, accessed on 14-04-2022.
- Graefe, U., Beylich, A. (2003): Critical values of soil acidification for annelid species and the decomposer community. In: *Newsletter on Enchytraeidae*, 8, 51–55.
- Graff, O. (1953): Die Regenwürmer Deutschlands. Schriftenreihe des Forschungsinstituts für Landwirtschaft, 7, 1–70.
- Griffiths, B. S., Römbke, J., Schmelz, R. M., Scheffczyk, A., Faber, J. H., Bloem, J., Pérès, G., Cluzeau, D., Chabbi, A., Suhadolc, M., Sousa, J. P., Martins da Silva, P., et al. (2016): Selecting cost effective and policy-relevant biological indicators for European monitoring of soil biodiversity and ecosystem function. In: *Ecological Indicators*, 69, 213–223.

- Groves, C. P. (2022): "Biogeographic region". In: *Encyclopedia Britannica*, <https://www.britannica.com/science/biogeographic-region>. Accessed on 26 April 2022.
- Guerra, C. A., Bardgett, R. D., Caon, L., Crowther, T. W., Delgado-Baquerizo, M., Montanarella, L., Navarro, L. M., Orgiazzi, A., Singh, B. K., et al. (2021): Tracking, targeting, and conserving soil biodiversity. In: *Science*, 371(6526): 239–241.
- Guisan, A., Thuiller, W. (2005): Predicting species distribution: offering more than simple habitat models. In: *Ecology Letters*, 8(9): 993–1009.
- Guisan, A., Thuiller, W., Zimmermann, N. E. (2017): *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge Univ. Press, 478 pp.
- Guisan, A., Zimmermann, N. E. (2000): Predictive habitat distribution models in ecology. In: *Ecological Modelling*, 135, 147–86.
- Hågvar, S. (1998): The relevance of the Rio-Convention on biodiversity to conserving the biodiversity of soils. In: *Applied Soil Ecology*, 9, 1–7.
- Harell, F., Califf, R., Pryor, D., Lee, K., Rosati, R. (1982): Evaluating the yield of medical tests. In: *Journal of the American Medical Association*, 247, 2543–2546.
- Hastie, T., Tibshirani, R. (1986): Generalized additive models. In: *Statistical Science*, 1, 297–318.
- Hedde, M., Bureau, F., Delporte, P., Cécillon, L., Decaëns, T. (2013): The effects of earthworm species on soil behaviour depend on land use. In: *Soil Biology & Biochemistry*, 65, 264–273.
- Hengl, T. (2007): *A Practical Guide to Geostatistical Mapping of Environmental Variables*. JRC Scientific and Technical Reports 38153 (EUR 22904 EN). Office for Official Publications of the European Communities, Luxembourg, 144 S.
- Hijmans, R. J., Elith, J. (2019) Spatial Distribution Models, spatial data science with R. <https://rspatial.org/sdm/SDM.pdf>.
- Hodson, M. E., Corstanje, R., Jones, D. T., Witton, J., Burton, V. J., Sloan, T., Eggleton, P. (2021): Earthworm distributions are not driven by measurable soil properties. Do they really indicate soil quality? In: *PLoS One*, 16(8): e0241945 (<https://doi.org/10.1371/journal.pone.0241945>).
- Hu, X. G., Jin, Y., Wang, X. R., Mao, J. F., Li, Y. (2015): Predicting impacts of future climate change on the distribution of the widespread conifer *Platycladus orientalis*. In: *PLoS One*, 10: e0132326 (<https://doi.org/10.1371/journal.pone.0132326>).
- Huber, S., Prokop, G., Arrouays, D., Banko, G., Bispo, A., Jones, R. J. A., Kibblewhite, M. G., Lexer, W., Möller, A., Rickson, R. J., Shishkov, T., Stephens, M., Toth, G., Van den Akker, J. J. H., Varallyay, G., Verheijen, F. G. A., Jones, A. R. (eds) (2008): *Environmental Assessment of Soil for Monitoring. Volume I: Indicators & Criteria*. EUR 23490 EN/1, Office for the Official Publications of the European Communities, Luxembourg, 339 S.
- Hurlbert, A. H., White, E. P. (2007): Ecological correlates of geographical range occupancy in North American birds. In: *Global Ecology and Biogeography*, 16, 764–773.
- IUCN (2012a): *Guidelines for Application of IUCN Red List Criteria at Regional and National Levels: Version 4.0*. IUCN, Gland, Switzerland and Cambridge, UK. Available at www.iucnredlist.org/technical-documents/categories-and-criteria.
- IUCN (2012b): *IUCN Red List Categories and Criteria: Version 3.1. 2nd edn*. IUCN, Gland, Switzerland and Cambridge, UK. Available at www.iucnredlist.org/technical-documents/categories-and-criteria.
- IUCN Standards and Petitions Committee (2022): *Guidelines for Using the IUCN Red List Categories and Criteria. Version 15*. <https://www.iucnredlist.org/documents/RedListGuidelines.pdf>.

- Jänsch, S., Steffens, L., Höfer, H., Horak, F., Roß-Nickoll, M., Russell, D., Toschki, A., Römbke J. (2013): State of knowledge of earthworm communities in German soils as a basis for biological soil quality assessment. In: *Soil Organism*, 85(3): 215–233.
- Jarnevich, C. S., Stohlgren, T. J., Kumar, S., Morisette, J. T., Holcombe, T. R. (2015): Caveats for correlative species distribution modelling. In: *Ecological Informatics*, 29(1): 6–15.
- Jiménez-Valverde, A. (2011): Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. In: *Global Ecology and Biogeography*, 21(4): 498-507 (<https://doi.org/10.1111/j.1466-8238.2011.00683>).
- Johnson, C. M., Johnson, L. B., Richard, C., Beasley, V. (2002): Predicting the occurrence of amphibians: An assessment of multiple-scale models. In: Scott, J. M., Heglund, P. J., Samson, F., Haufler, J., Morrison, M., Raphael, M., Wall, B. (Eds.): *Predicting species occurrences: Issues of accuracy and scale*. Island Press, Covelo, CA, S. 157–170.
- Kalu, S., Koirala, M., Khadaka, R. J. (2015): Earthworm population in relation to different land use and soil characteristics. In: *Journal of Ecology and the Natural Environment*, 7(5): 124–131.
- Kaniarska, R.; Jadůdová, J.; Makovníková, J.; Kizeková, M. (2016): Assessment of Relationships between Earthworms and Soil Abiotic and Biotic Factors as a Tool in Sustainable Agricultural. In: *Sustainability*, 8, 906 (<https://doi.org/10.3390/su8090906>).
- Kaufman-Boll, A., Lazar, S., Rinklebe, J. (2011): Auswertung der Veränderungen des Bodenzustands für Boden-Dauerbeobachtungsflächen (BDF) und Validierung räumlicher Trends unter Einbeziehung anderer Messnetze. Teil A: Methoden-Code und Umgang mit Verfahrenswechseln. UBA Texte 89/2011, Umweltbundesamt, Berlin, 330 S.
- Keith, D. A., Akçakaya, H. R., Murray, N. J. (2018): Scaling range sizes to threats for robust predictions of risks to biodiversity. In: *Conservation Biology*, 32, 322–332.
- Krück, S. (2018): Bildatlas zur Regenwurmbestimmung. Natur+Text, Rangsdorf. 196 S.
- Krück, S., Joschko, M., Schultz-Sternberg, R., Kroschewski, B., Tessmann, J. (2006): A classification scheme for earthworm populations (Lumbricidae) in cultivated agricultural soils in Brandenburg, Germany. In: *J. Plant Nutr. Soil Sci.*, 169, 651–660.
- Kuhn, M. (2008): Building Predictive Models in R Using the caret Package. In: *Journal of Statistical Software*, 28(5): 1–26.
- Kuhn, M. (2022): Package ‘caret’: Classification and Regression Training. <https://cran.r-project.org/web/packages/caret/index.html>.
- Kumar, S., Stohlgren, T. J. (2009): Maxent modelling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola*. In: *New Caledonia Journal of Ecology and Natural Environment*, 1, 94–98.
- Lavelle, P. (1978): Les vers de terre de la savane de lamto (Côte d'Ivoire) peuplements, population et fonctions dans l'écosystème. Dissertation, Université Paris VI/ENS, 301 S.
- Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.-P. (2006): Soil invertebrates and ecosystem services. In: *European Journal of Soil Biology*, 42: S3–S15.
- Lavelle, P., Spain, V. A. (2005): *Soil Ecology*. Springer Science+Business Media, Dordrecht. 682 S.
- Lê, S., Josse, J., Husson, F. (2008): FactoMineR: A Package for Multivariate Analysis. In: *Journal of Statistical Software*, 25(1): 1–18.
- Lee, K. E. (1985): *Earthworms: their Ecology and Relationships with Soils and Land Use*. Academic Press, Ann Arbor. 411 S.

- Lehmitz, R., Römbke, J., Graefe, U., Beylich, A., Krück, S. (2016): Rote Liste und Gesamtartenliste der Regenwürmer (Lumbricidae et Criodrillidae) Deutschlands. In: Naturschutz und Biologische Vielfalt, 70(4): 565–590.
- Li, X., Wang, Y. L. (2013): Applying various algorithms for species distribution modelling. In: Integrative Zoology, 8, 124–135.
- Liaw, A., Wiener, M. (2002): Classification and Regression by randomForest. In: R News, 2(3), 18–22. <https://CRAN.R-project.org/doc/Rnews/>.
- Liu, J.; Cosman, P. C.; Rao, B. D. (2018): Robust Linear Regression via l0 Regularization. In: IEEE Transactions on Signal Processing, 66(3), 698–713.
- Maes, J., Ego, B., Willemen, L., Liquete, C., et al. (2012): Mapping ecosystem services for policy support and decision making in the European Union. In: Ecosyst. Services, 1, 31–39.
- Magurran, A. E. (2004): Measuring Biological Diversity. Blackwell Publ., Malden, Oxford, Carlton. 256 S.
- Manel, S., Ceri Williams, H., Ormerod, S. J. (2001): Evaluating presence-absence models in ecology: The need to account for prevalence. In: Journal of Applied Ecology, 38, 921–931.
- Mangiafico, S. S. (2016): Summary and Analysis of Extension Program Evaluation in R, version 1.18.8. (<https://rcompanion.org/handbook/>).
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., Thuiller, W. (2008): Evaluation of consensus methods in predictive species distribution modelling. In: Divers. Distrib., 15, 56–69.
- McCullagh, P., Nelder, J. A. (1989): Generalized Linear Models. Chapman and Hall, London. 532 S.
- Mi, C., Huettman, F., Guo, Y., Wen, L. (2017): Why choose random forest to predict rare species distribution with few samples in large undersampled area? Three Asian crane species models provide supporting evidence. In: PeerJ, 5:e2849 (DOI:10.7717/peerj.2849).
- Mod, H. K., Scherrer, D., Luoto, M., Guisan, A. (2016): What we use is not what we know: environmental predictors in plant distribution models. In: Journal of Vegetation Science, 27, 1308–1322.
- Ockleford, C., Adriaanse, P., Berny, P., Brock, T., et al. (2017): Scientific Opinion addressing the state of the science on risk assessment of plant protection products for in-soil organisms. EFSA PPR Panel. EFSA Journal 15(2), doi:10.2903/j.efsa.2017.4690.
- Orgiazzi, A., Panagos, P., Yigini, Y., Dunbar, M. B., Gardi, C., Montanarella, L., Ballabio, C. (2016): A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. In: Science of the Total Environment, 545–546, 11–20.
- Pebesma, E. J., Bivand, R. S. (2005): “Classes and methods for spatial data in R.” R News, 5(2): 9–13 (<https://CRAN.R-project.org/doc/Rnews/>).
- Pérès, G., Cluzeau, D., Curmi, P., Hallaire, V. (1998): Earthworm activity and soil structure changes due to organic enrichments in vineyard systems. In: Biology and Fertility of Soils, 27, 417–424.
- Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., et al. (2019): Global distribution of earthworm diversity. In: Science, 366(6464), 480–485.
- Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., Ferrier, S. (2008): Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. In: Ecological Applications, 19, 181–197.
- Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., Rutgers, M. (2012): Soil biodiversity, biological indicators and soil ecosystem services — an overview of European approaches. In: Current Opinion in Environmental Sustainability, 4, 529–538.

- Pulliam, H. R. (2000): On the relationship between niche and distribution. In: *Ecology Letters*, 3, 349–361.
- R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (<https://www.R-project.org/>).
- Rajwar, N., Singh, V., Bhatt, S., Singh Bisht, S. (2022): Earthworm population dynamics in three different land use systems along an altitudinal gradient (208–2609 m asl) in Kumaun Himalayas, India. In: *Tropical Ecology*, 63, 134–140.
- Reuter, H. I., Lado, L. R., Hengl, T., Montanarella, L. (2008): Continental-scale digital soil mapping using European soil profile data: soil pH. In: *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, 19, 91–102.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., Dorman, C. F. (2017): Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. In: *Ecography*, 40, 913–929.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., Müller, M. (2011): pROC: an open-source package for R and S+ to analyze and compare ROC curves. In: *BMC Bioinformatics*, 12, 77 (<https://doi.org/10.1186/1471-2105-12-77>).
- Römbke, J., Blick, T., Dorow, W. H. O. (2017): *Allolobophoridella eiseni* (Lumbricidae), a truly arboreal earthworm in the temperate region of Central Europe. In: *Soil Organisms*, 89(2): 75–84.
- Römbke, J., Dorow, W. H. O., Jänsch, S. (2018): Distribution and diversity of earthworms (Lumbricidae) in Hesse (Central Germany): current knowledge. In: *Soil Organisms*, 90(3): 171–185.
- Römbke, J., Dreher, P., Beck, L., Hammel, W., Hund, K., Knoche, H., Kördel, W., Kratz, W., Moser, T., Pieper, S., Ruf, A., Spelda, J., Woas, S. (2000): Bodenbiologische Bodengüte-Klassen. Umweltbundesamt, Berlin. 276 S.
- Römbke, J., Dreher, P., Beck, L., Hund-Rinke, K., Jänsch, S., Kratz, W., Pieper, S., Ruf, A., Spelda, J., Woas, S. (2002): Entwicklung von bodenbiologischen Bodengüteklassen für Acker- und Grünlandstandort. UBA Texte 20/02. Umweltbundesamt, Berlin. 273 S.
- Rutgers, M., Mulder, C., Schouten, A. J. (eds.) (2008): Soil ecosystem profiling in the Netherlands with ten references for biological soil quality. RIVM Report 607604009/2008. RIVM, Bilthoven (NL). 86 S.
- Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A. M., Neilson, R., Boag, B., Schmidt, O., et al. (2016): Mapping earthworm communities in Europe: In: *Applied Soil Ecology*, 97, 98–111.
- Rutgers, M., Schouten, A. J., Bloem, J., van Eekeren, N., de Goede, R. G. M., Jagers op Akkerhuis, G. A. J. M., van der Wal, A., Mulder, C., Brussaard, L., Breure, A. M. (2009): Biological measurements in a nationwide soil monitoring network. In: *European J. Soil Science*, 60, 820–832.
- Salako, G., Chandalin, B., Aliyu, M. B., Sawyer, H. (2015): Modeling the suitability index of selected conifers on Mambilla Plateau Taraba State, Nigeria: Implication on planted forest. In: *International Journal of Agroforestry Remote Sensing and GIS*, 1(1): 1–9.
- Salako, G., Oyebanji, O. O., Olagunju, T. E., Howe, G. T. (2020): Potential impact of climate change on the distribution of some selected legumes in Cameroon and adjoining Nigeria border. In: *African Journal of Ecology*, 00, 1–17 (<https://doi.org/10.1111/aje.1291>).
- Schmid, M., Wright, M. N., Ziegler, A. (2016): On the use of Harrell's C for clinical risk prediction via random survival forests. arXiv 1507:03092v2 (<https://arxiv.org/pdf/1507.03092.pdf>).
- Sherlock, E. (2018): Key to the earthworms of the UK and Ireland (2nd Edn.). FSC Publ., Shrewsbury. 92 S.
- Sheth, S. N., Morueta-Holme, N., Angert, A. L. (2020): Determinants of geographic range size in plants. In: *New Phytologist*, 226, 650–665.

- Sillero, N., Márcia Barbosa, A. (2021): Common mistakes in ecological niche models. In: International Journal of Geographical Information Science, 35(2): 213–226.
- Sims, R. W., Gerard, B. M. (1999): Earthworms. FSC Publ., Shrewsbury. 169 S.
- Soil Security Laboratory (2017): Continuous soil attribute modeling and mapping: Multiple linear regression. 10 S.
- Spurgeon, D. J., Keith, A. M., Schmidt, O., Lammertsma, D. R., Faber, J. H. (2013): Land-use and land-management change: relationships with earthworm and fungi communities and soil structural properties. In: BMC Ecol, 13, 46 (<https://doi.org/10.1186/1472-6785-13-46>).
- Suthar, S. (2009): Earthworm communities as bio-indicator of arable land management practices: A case study in semiarid region of India. In: Ecol. Indic., 9, 588–594.
- Thuiller, W., Araujo, M. B., Lavorel, S. (2004): Do we need land-cover data to model species distributions in Europe? In: Journal of Biogeography, 31, 353–361.
- Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F. (2021): Package ‘biomod2’: Ensemble Platform for Species Distribution Modeling. <https://cran.r-project.org/web/packages/biomod2/index.html>.
- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M. B. (2008): BIOMOD - a platform for ensemble forecasting of species distributions. In: Ecography, 32, 369–373.
- Toschki, A., Burkhardt, U., Haase, H., Höfer, H., Jänsch, S., Oellers, J., Römbke, J., Roß-Nickoll, M., Salomon, J.-A., Schmelz, R. M., Scholz-Starke, B., Russell, D. (2021): Die Edaphobase-Länderstudie: Synökologische Untersuchungen von Bodenorganismen in einem Biotop- und Standortgradienten in Deutschland 2014–2018. Peckiana 14. Senckenberg, Görlitz. 367 S.
- Turbé, A., de Toni, A., Benito, P., Lavelle, P., Lavelle, P., Ruiz Camacho, N., van der Putten, W. H., Labouze, E., Mudgal, S. (2010) Soil biodiversity: functions, threats and tools for policy makers. Report to the European DG Environment, 249 S.
- UBA (Umweltbundesamt) (2002): Boden-Dauerbeobachtung in Deutschland. Ergebnisse aus den Ländern. UBA Texte 66/02, Umweltbundesamt, Berlin, 146 S.
- Urbani, F., Dalessandro, P., Biondi, M. (2017): Using Maximum Entropy Modeling (MaxEnt) to predict future trends in the distribution of high altitude endemic insects in response to climate change. In: Bulletin of Insectology, 70, 189–200.
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., Guillera-Arroita, G. (2021): Modelling species presence-only data with random forests. In: Ecography, 44(12): 1731–1742.
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J. (2022): Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. In: Ecological Monographs, 92(1): e0148 (<https://doi.org/10.1002/ecm.1486>).
- van Leeuwen, J. P., Saby, N. P. A., Jones, A., Louwagie, G., Micheli, E., Rutgers, M., Schulte, R. P. O., Spiegel, H., Toth, G., Creamer, R. E. (2017): Gap assessment in current soil monitoring networks across Europe for measuring soil functions. In: Environmental Research Letters, 12, 124007 (<https://doi.org/10.1088/1748-9326/aa9c5c>).
- Verberk, W. (2011): Explaining General Patterns in Species Abundance and Distributions. In: Nature Education Knowledge, 3(10), 38.
- Wang, R., Li, Q., He, S., Liu, Y., Wang, M., Jiang, G. (2018). Modeling and mapping the current and future distribution of *Pseudomonas syringae* pv. *actinidiae* under climate change in China. In: PLoS One, 13, e0192153 (<https://doi.org/10.1371/journal.pone.0192153>).

Wardle, D. A. (2002): Communities and Ecosystems – Linking the Aboveground and Belowground Components. Princeton Univ. Press, Princeton Oxford. 392 S.

Weeks, J. M. (1998): A Demonstration of the Feasibility of SOILPACS. – Environmental Agency, London, 180 S.

Wickham, H. (2016): ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 213 S.

Wieger Wamelink, G. W., Walvoort, D. J. J., Sanders, M. E., Meeuwsen, H. A. M., Wegman, R. M. A., Pouwels, R., Knotters, M. (2019): Prediction of soil pH patterns in nature areas on a national scale. In: Applied Vegetation Science, 22(2), 189-199.

Yiu T. (2021): Understanding Random forest. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., *et al.* (2020): A standard protocol for reporting species distribution. In: Ecography, 43, 1261–1277.

A Anhang

Tabelle 12: Datentabelle zu Abbildung 28: Vorkommenswahrscheinlichkeiten der Regenwurmarten in zehn Habitattypen.

	Meer	Küsten	Auen	Moore	Grünland	Gebüsch	Wälder	Spärl. Veg.	Acker	Urban/ Bergbaufol.	Mittelwert
<i>A. caliginosa</i>	0,05	0,20	0,51	0,15	0,55	0,50	0,55	0,34	0,52	0,56	0,39
<i>A. chlorotica</i>	0,03	0,05	0,04	0,10	0,60	0,10	0,30	0,10	0,60	0,30	0,22
<i>A. eiseni</i>	0,01	0,20	0,20	0,11	0,48	0,12	0,65	0,04	0,45	0,40	0,27
<i>A. limicola</i>	0,34	0,20	0,55	0,46	0,50	0,20	0,58	0,30	0,52	0,10	0,38
<i>A. longa</i>	0,05	0,12	0,07	0,15	0,51	0,38	0,52	0,30	0,56	0,41	0,31
<i>A. rosea</i>	0,03	0,02	0,40	0,30	0,58	0,01	0,60	0,12	0,59	0,45	0,31
<i>D. attemsi</i>	0,10	0,06	0,13	0,35	0,30	0,25	0,45	0,24	0,55	0,10	0,25
<i>D. octaedra</i>	0,10	0,20	0,25	0,20	0,26	0,20	0,64	0,43	0,42	0,34	0,30
<i>L. badensis</i>	0,02	0,02	0,05	0,02	0,30	0,00	0,50	0,10	0,20	0,02	0,12
<i>L. castaneus</i>	0,20	0,00	0,50	0,01	0,35	0,01	0,59	0,11	0,48	0,44	0,27
<i>L. rubellus</i>	0,30	0,02	0,35	0,01	0,56	0,10	0,60	0,04	0,48	0,34	0,28
<i>L. terrestris</i>	0,01	0,04	0,35	0,12	0,59	0,03	0,62	0,11	0,20	0,25	0,23
Summe	1,24	1,13	3,40	1,98	5,58	1,90	6,60	2,23	5,57	3,71	

Tabelle 13: Datentabelle zu Abbildung 29: Vorkommenswahrscheinlichkeiten der Regenwurmarten größer als der Schwellenwert 0,2 in zehn Habitattypen.

	<i>A. caliginosa</i>	<i>A. chlorotica</i>	<i>A. eiseni</i>	<i>A. limicola</i>	<i>A. longa</i>	<i>A. rosea</i>	<i>D. attemsi</i>	<i>D. octaedra</i>	<i>L. badensis</i>	<i>L. castaneus</i>	<i>L. rubellus</i>	<i>L. terrestris</i>
Meer					0,34						0,20	0,30
Küsten	0,20			0,20	0,20							
Auen	0,51			0,20	0,55	0,40		0,25		0,50	0,35	0,35
Moore					0,46	0,30	0,35					
Grünland	0,55	0,60	0,48	0,50	0,51	0,58	0,30	0,26	0,30	0,35	0,56	0,59
Gebüsch		0,50			0,20	0,38	0,25					
Wälder	0,55	0,30	0,65	0,58	0,52	0,60	0,45	0,64	0,50	0,59	0,60	0,62
Spärliche Vegetation			0,34		0,30	0,30	0,24	0,43				
Acker	0,52	0,60	0,45		0,56	0,59	0,55	0,34		0,48	0,48	0,20
Urban/Bergbaufolge		0,56	0,30	0,40		0,41	0,45	0,34		0,44	0,34	0,25
	3,73	1,80	2,38	3,13	2,68	2,92	2,14	2,26	0,80	2,56	2,63	2,01