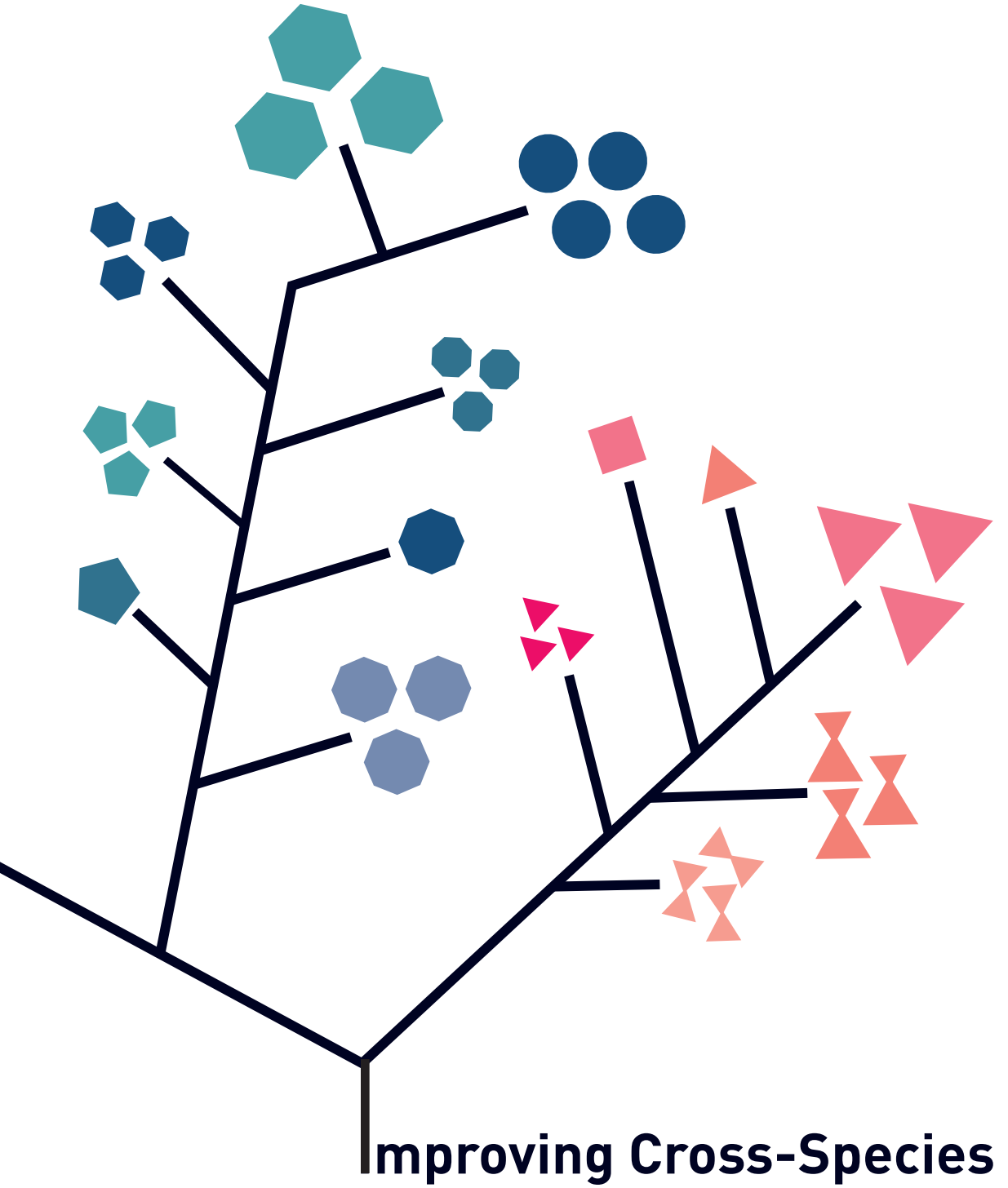


Improving Cross-Species Extrapolation of Chemical Sensitivity

Sanne van den Berg

Improving cross-species extrapolation of chemical sensitivity

Sanne van den Berg

Thesis committee

Promotor

Prof. Dr Paul J. van den Brink

Personal chair, Chemical Stress Ecology

Wageningen University & Research

Co-promotor

Prof. Dr Frederik de Laender

Laboratory of Environmental Ecosystems Ecology

Research Unit in Environmental and Evolutionary Biology

University of Namur, Belgium

Other members

Prof. Dr Lourens Poorter, Wageningen University & Research

Dr Michiel Daam, Nova University, Lisbon, Portugal

Prof. Dr Nico M. van Straalen, VU University Amsterdam

Prof. Dr Leo Posthuma, National Institute for Public Health and the Environment (RIVM), Bilthoven

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Improving cross-species extrapolation of chemical sensitivity

Sanne van den Berg

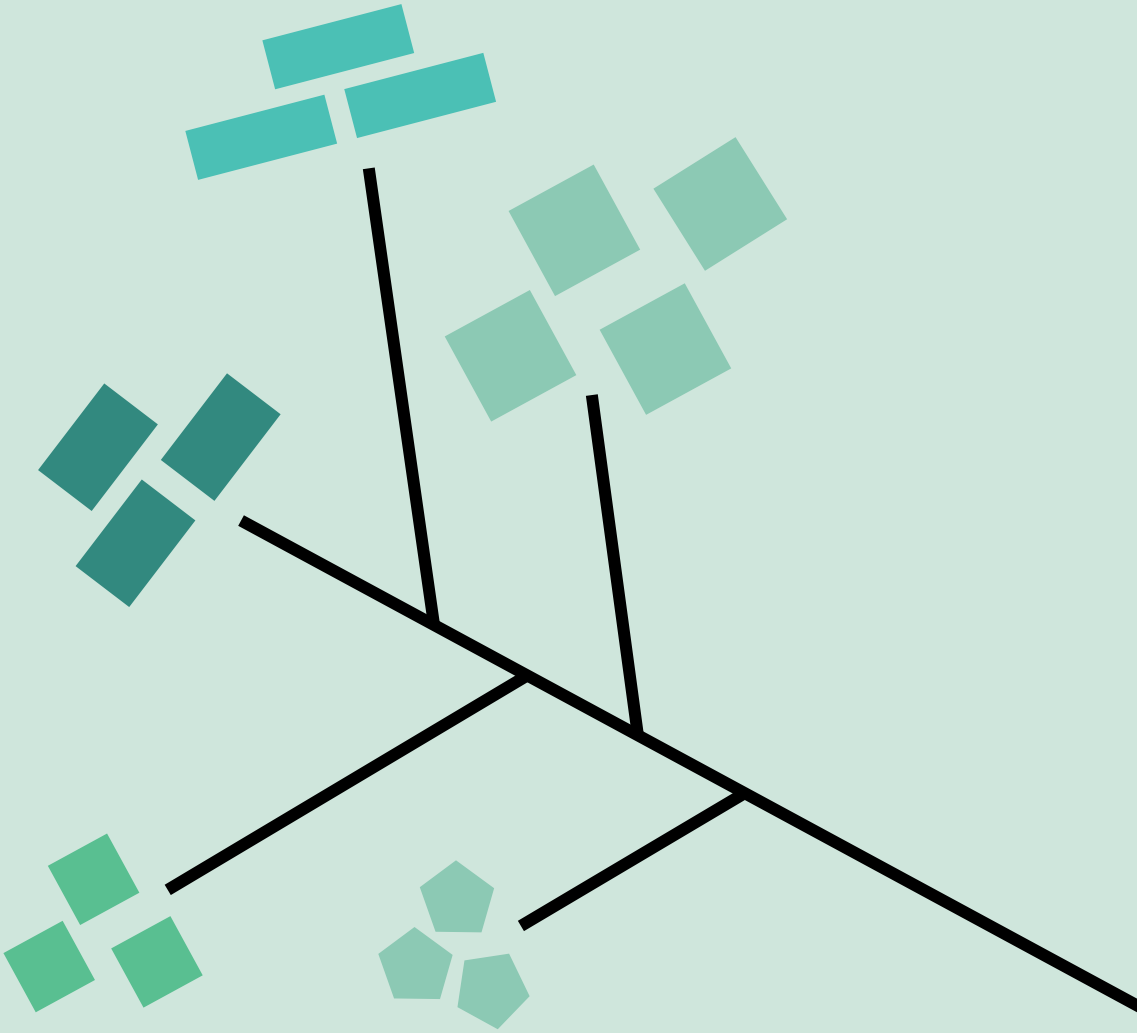
Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. de Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 12 June 2020
at 16:00 in the Aula

Sanne van den Berg
Improving Cross-Species Extrapolation of Chemical Sensitivity
150 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summary in English
ISBN: 978-94-6395-392-4
DOI: https://doi.org/10.18174/520570

Contents



CHAPTER 1

Introduction

1.1 Why do we need cross-species extrapolation?

Chemicals provide substantial benefits to society, but their widespread use in industry, agriculture, and homes has led in some cases to pollution of land, water, and air. Since the current world population is expected to increase from 7.7 billion people in 2019 to 9.7 billion in 2050 (United Nations 2019), we can expect an increase in the impact of chemicals on natural ecosystems, unless we manage to reduce the adverse effects associated with these chemicals. The adaptation of precautionary approaches to identify the risks of hazardous chemicals in the environment is a requirement when we want to avoid an adverse impact of chemicals on the ecosystems surrounding us. However, traditional risk assessments lack mechanistic insight, rely on in vivo exposure, and apply arbitrary assessment factors to assess chemical safety (Brock, Arts et al. 2006). Greater mechanistic insights and the application of a systems-based approach can be achieved by the development and use of predictive models. Such models would enable risk assessors to extrapolate from laboratory studies on single organisms to assemblages in the field. In addition, such insights could help to identify environmental hazards earlier in chemical development, predict potential impacts of chemicals on specific taxonomic groups and assess interactions with other stressors on ecosystem function and ecosystem services. Therefore, this thesis focuses on developing and applying predictive methodologies for the extrapolation of chemical sensitivity across species.

An environmental risk assessment (ERA) is used to obtain insight into the potential effects of a chemical on an ecosystem, and traditionally entails two different phases, the exposure and the effect assessment. The exposure assessment, when applied in the prospective risk assessment of chemicals required before market authorization, depends on chemical concentration dynamics calculated for different environmental compartments (e.g. soil, water, sediments) using mathematical models. When applied in the retrospective risk assessment of chemicals, aimed to identify the causes of adverse effects that have already occurred, the exposure assessment can be based on measured concentrations. Eventually, the exposure assessment results in a predicted environmental concentration (PEC). In the effect assessment, a regulatory acceptable concentration is determined. This acceptable concentration depends on which effects on organisms, populations, or ecosystems functions or services are tolerated under the protection goals at force under the legal framework (e.g. the Water Framework Directive, European Commission 2000). This regulatory acceptable concentration is, for aquatic ecosystems, traditionally based on effects found for the organism groups: algae, fish and invertebrates, and results in a predicted no-effect concentration (PNEC, Brock, Arts et al. 2006). The ratio between the PEC and the PNEC is then combined with an assessment factor to account for extrapolation from the field to the lab, from one to many species (including interactions), and from short- to long-term exposure, to result in the risk quotient. This risk quotient then forms the foundation of chemical management decisions.

This traditional ERA approach remains rather general, as it results in one value that is then applied to any system, anywhere, anytime, and thereby ignores differences in the sensitivity of species over space and time. However, since it is impossible to experimentally determine the sensitivity of all the species present in any ecosystem to all chemicals to which they can possibly be exposed and under all environmental conditions they might be exposed under, we rely on cross-species extrapolation of chemical sensitivity. To do this, however, requires on the one hand a mechanistic understanding of what determines sensitivity, and on the other hand an understanding of which factors contribute to the different chemical response of one species compared to another.

Concerning the mechanistic understanding of sensitivity, a large branch of ecotoxicological research has investigated the mechanisms through which chemicals cause damage at a cellular level. This led to the description of so-called modes or mechanisms of action (e.g. Escher and Hermens 2002, Enoch, Hewitt et al. 2008). These modes of action classify chemicals according to the mechanisms through which they cause a molecular response that eventually results in damage at a cellular level (Escher and Hermens 2002). The advantage of these modes of action is that these mechanisms are often conserved over large taxonomic groups. This makes them perfect to study differences in species sensitivity, because if the mechanism of toxicity is the same, then what causes the differences in the response of multiple species?

This brings us to the second aspect, focusing on which factors help to discriminate one species from another in their response to a chemical stressor. For this, a logic start is to describe what makes one species different from another species. At the most basic biological level, species differ from each other in their DNA. Studies at this level have shown that the simple presence or absence of the molecular target of the chemical determines that one species is susceptible to a chemical, whilst another is not (e.g. Kim and Lee 2013). Of course, more complex mechanisms complicate effect processes occurring at this level, for instance, the existence of orthologs, homologs and analogues (Moreno-Hagelsieb and Latimer 2007). Nevertheless, the sequence similarity of a known molecular target and its orthologs can be used to determine if and to which extent a species is potentially sensitive towards a chemical (LaLone, Villeneuve et al. 2013). These more complex and accurate pathways of how molecular responses to chemical stress result in effects at the individual level were researched in great detail for several model species (i.e. adverse outcome pathways, Ankley, Bennett et al. 2010). However, for most species, these pathways remain unknown, and we rely on simple mode of action classifications for information on the mechanism of toxicity.

As we know from nature, even very closely related species with a very similar genome, can look very different. Consider the difference in the appearances of a chimpanzee and a human, with a staggering amount of 95% similarity in our genomes (Britten 2002). Therefore, what else characterizes one species from another? The answer to that question is their traits. A trait is a phenotypic or ecological character of an organism at individual or population level, and

describes the physical characteristics, ecological niche and functional role of a species within the ecosystem (Baird, Rubach et al. 2008). Trait-based approaches have long been included in bioassessment approaches (most famously, in the River Continuum Concept, Vannote, Minshall et al. 1980), adding mechanistic and diagnostic knowledge to species occurrence patterns found in the field (well-explained in Culp, Armanini et al. 2011). In the last decade, trait-based approaches have also been introduced in ecotoxicology (Baird, Rubach et al. 2008), due to the growing realization that a solely taxonomic approach can limit our understanding of how a system responds to stress. They have proven that the use of traits can improve our understanding of stress responses, and may help to facilitate the prediction of chemical stress (Rubach, Baird et al. 2012).

As described so far in this chapter, effects occur at sub-cellular level, whilst most measurements of species sensitivity occur at the level of the whole individual (e.g. by a measure of reduced fitness i.e. higher mortality rate, lower reproduction rate). So how can we make these two worlds meet? In the approaches revealed in this thesis, we combine mechanistic information on chemical toxicity with ecological knowledge at the individual and community scale, in order to make a case for a more regional, fit-for-purpose ERA.

1.2 Objectives

The overall aim of this thesis is to support the development of models describing species sensitivity towards chemical stressors. This includes unravelling decisions that might be of importance in the modelling process, obtaining a better mechanistic understanding of differences in species sensitivity, and providing recommendations for applying modelling approaches within or across different taxonomic groups in freshwater ecosystems all over the world.

To accomplish this global applicability, this thesis primarily focuses on the use of trait-based approaches. Besides that traits add mechanistic and diagnostic knowledge (Rubach, Ashauer et al. 2011), they have additional advantages in that they are transferable across geographies (Usseglio-Polatera, Bournaud et al. 2000, Bellwood, Wainwright et al. 2002, Van den Brink, Alexander et al. 2011), and can be easily translated from taxonomic analyses priory preformed (Van den Brink, Alexander et al. 2011). However, besides using trait-based approaches solely, we also studied the potential of hybrid approaches in which traits are combined with other predictors (e.g. relatedness). Finally, we explored the use of alternative descriptors of sensitivity.

1.3 Thesis outline

In **Chapter 2**, we report about a new predictive modelling approach for potential use in environmental risk assessment (ERA). This new approach constructs macroinvertebrate sensitivity rankings, and subsequently, predictive trait-based models for a set of pre-defined modes of action. Each model reveals interesting taxonomic patterns of species sensitivity, as

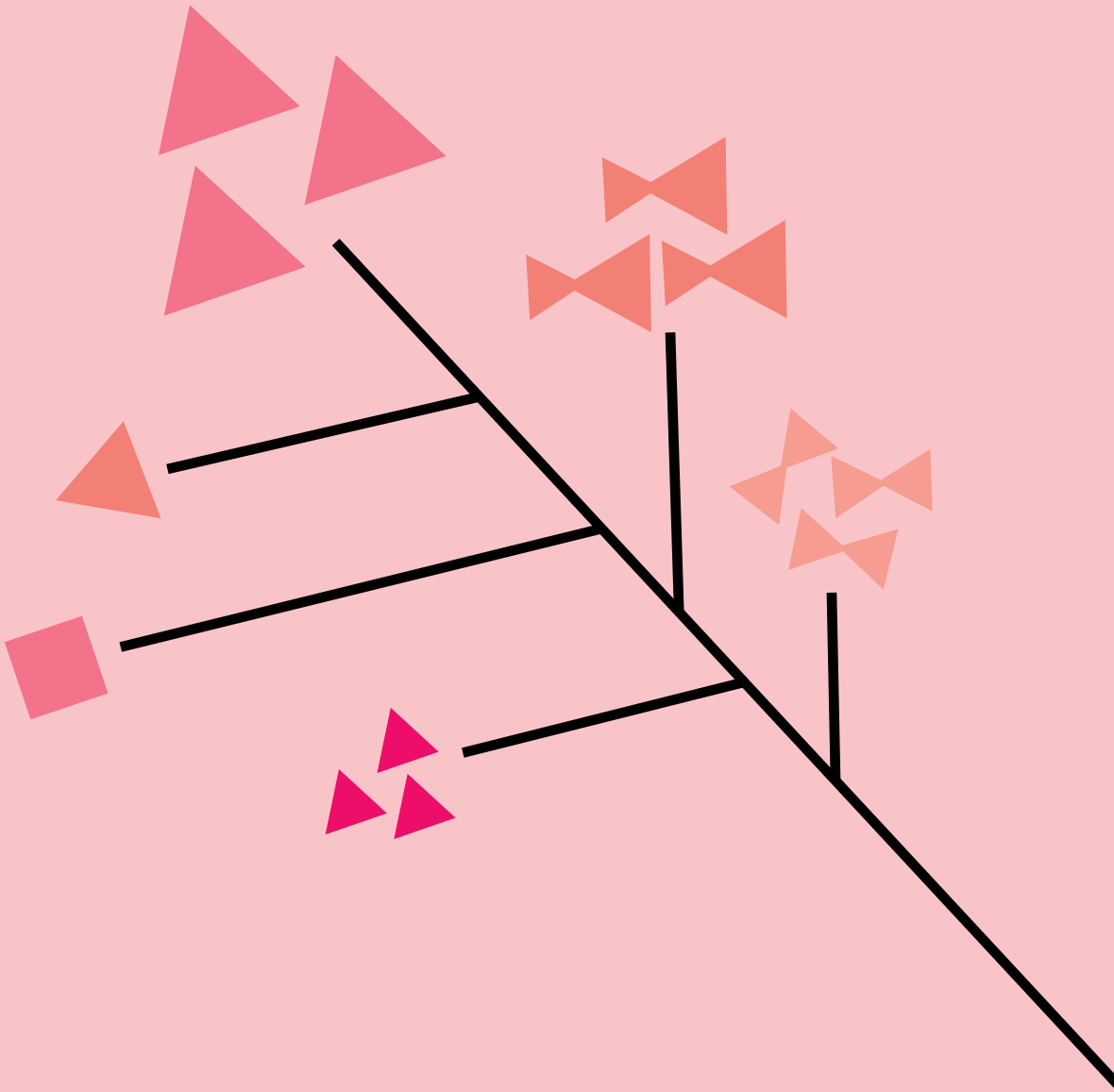
well as an accurate picture of which species-chemical combinations still lack sufficient data. Additionally, we demonstrate that there is not one (set of) species which is most sensitive to all chemicals.

In **Chapter 3,** we develop the models of Chapter 2 into hybrid models by the addition of taxonomic relatedness. Subsequently, we apply the models at two different spatial scales, continental and national, revealing spatial patterns in species sensitivity towards chemical stressors. We conduct the first trait-based chemical sensitivity assessment of freshwater macroinvertebrate assemblages, and test the influence of spatial scale on sensitivity patterns. Finally, we provide important insights and recommendations for the improved application of the developed method to invertebrates, an organism group which remains underrepresented in global conservation priorities.

Chapter 4 proves that the use of a more mechanistic description of sensitivity improves the predictive power of cross-species extrapolation models. We demonstrate this using the quantitative mechanistic toxicokinetic (TK)- toxicodynamic (TD) models of the General Unified Threshold models of Survival (GUTS) framework. These models link external exposure and survival effects by describing dynamically the process of TK (uptake, biotransformation, and elimination) and TD (damage/hazard, internal recovery and thresholds). We aim to predict these TKTD parameters using species traits, and that traits-predicted TKTD parameters can subsequently be used to predict classical sensitivity endpoints such as LC50.

Chapter 5 gives an overview of currently existing cross-species extrapolation methods, dividing them into descriptive groups, and provides an overview of the trade-offs that exist between them. We also give advice on crucial modelling decisions, and demonstrate that these decisions are highly dependent on the taxonomic group and the mode of action of the chemical under study.

Finally, **Chapter 6** provides a synthesis and general discussion, summarizing all important findings of this thesis, and puts them in a broader perspective. The final chapter also provides recommendations for the incorporation of cross-species extrapolation models into a prospective ERA framework.



CHAPTER 2

Modeling the sensitivity of aquatic macroinvertebrates to chemicals using traits

Sanne J. P. Van den Berg, Hans Baveco, Emma Butler, Frederik De Laender, Andreas Focks, Antonio Franco, Cecilie Rendal, Paul J. Van den Brink

Abstract

In this study, a trait-based macroinvertebrate sensitivity modelling tool is presented that provides two main outcomes: (1) it constructs a macroinvertebrate sensitivity ranking and, subsequently, a predictive trait model for each one of a diverse set of pre-defined Modes of Action (MOAs) and (2) it reveals data gaps and restrictions, helping with the direction of future research. Besides revealing taxonomic patterns of species sensitivity, we find that there was not one genus, family or class which was most sensitive to all MOAs, and that common test taxa were often not the most sensitive at all. Traits like life cycle duration and feeding mode were identified as important in explaining species sensitivity. For 71% of the species, no or incomplete traits data were available, making the lack of trait data the main obstacle in model construction. Research focus should therefore be on completing trait databases, and enhancing them with finer morphological traits, focusing on the toxicodynamics of the chemical (e.g. target site distribution). Further improved sensitivity models can help with the creation of ecological scenarios by predicting the sensitivity of untested species. Through this development, our approach can help reduce animal testing and contribute towards a new predictive ecotoxicology framework.

This chapter is based on the paper: Sanne J. P. van den Berg, Hans Baveco, Emma Butler, Frederik De Laender, Andreas Focks, Antonio Franco, Cecilie Rendal, Paul J. Van den Brink (2019). Modelling the sensitivity of aquatic macroinvertebrates to chemicals using traits. Environmental Science and Technology, 53 (10), 6025-6034.

2.1 Introduction

In the environmental risk assessment (ERA) of chemicals it is essential to determine the environmental threshold concentration below which ecosystem structure and functioning experience no adverse impacts. In order to set this threshold, a key challenge in ERA remains the extrapolation of effects of toxicants found for a limited number of standard test species to many additional species. Ecosystems are generally populated by hundreds to thousands of species, and each species has the potential to show a different sensitivity towards one of the hundreds or even thousands of different chemical compounds that can be present in our ecosystems (Guénard, Ohe et al. 2011, Guénard, von der Ohe et al. 2014). Experimental testing of this innumerable amount of species-chemical combinations, plus any possible environmentally realistic mixture of those chemicals, is impossible. We therefore need to improve current modelling approaches and make them flexible for application to any geographic region and any set of abiotic conditions.

Traditional approaches trying to incorporate species diversity into risk assessment include the application of uncertainty factors (Chapman, Fairbrother et al. 1998) and the fitting of species sensitivity distributions (SSDs, Posthuma, Suter II et al. 2001) to available toxicity data. While both methods are extensively used (and frequently combined), they are aimed at being protective rather than predictive and as such, still maintain large uncertainty due to both a limited knowledge of the mechanisms underlying species sensitivity and a lack of taxonomic diversity. This lack of taxonomic diversity is especially true for uncertainty factors, but also holds for SSDs. Regulatory frameworks require SSDs to contain 10 to 15 species in total (Belanger, Barron et al. 2017), but divided over the different organism groups (e.g. fish, crustaceans, algae), this results in only 1 or 2 organisms from each organism group, which are often comprised out of the same set of standard test species. For the extrapolation across chemicals, Quantitative Structure-Activity Relationships (QSARs) are commonly used (Donkin 2009). QSARs use chemical characteristics to predict the toxicity of many chemicals for a certain species. Due to their large demand for experimental toxicity data, however, QSAR-models are often built only for specific standard test species like *Daphnia magna* and *Oncorhynchus mykiss* and, therefore, fail to account for the large species diversity of real ecosystems. In the last decade, trait-based approaches have been introduced to overcome this lack of realism (Baird and Van den Brink 2007, Rubach, Ashauer et al. 2011). These approaches incorporate more ecological realism into ERA by considering traits to provide a clear mechanistic link between exposure and effects, making it possible to extrapolate species sensitivities over chemicals acting by the same Mode Of Action (MOA).

After the introduction of trait-based approaches as a potential tool in ERA around a decade ago (Liess and Von der Ohe 2005, Baird and Van den Brink 2007, Buchwalter, Cain et al. 2007, Baird, Rubach et al. 2008), they have been rapidly evolving (Rubach, Baird et al. 2010, Liess and Beketov 2011, Rubach, Ashauer et al. 2011, Segner 2011, Van den Brink, Alexander et al.

2011, Ippolito, Todeschini et al. 2012, Rico and Van den Brink 2015, Baert, De Laender et al. 2017). Baird and Van den Brink (2007) were among the first to use biological traits to predict species sensitivity. They performed a Principal Components Analysis (PCA, Ter Braak 1995) on a species-by-substance matrix (12 macroinvertebrate species, 15 chemicals covering several MOAs), where they introduced a species-by-traits matrix as a set of nominal, passive explanatory variables. They found that up to 71% of the variability in the sensitivity could be explained with only four species traits. In a later study, Rubach, Baird and Van den Brink (2010) developed the approach further, now using single and multiple linear regression instead of PCA, and dividing the chemicals into groups according to their MOA. MOA has proven to be a strong determinant of species sensitivity and is, therefore, seen as a promising alternative to chemical class-based predictive toxicity modelling (Barron, Lilavois et al. 2015, Kienzler, Barron et al. 2017). Rubach , Baird and Van den Brink (2010) defined the Mode Specific Sensitivity (MSS) value of each species as the average relative sensitivity of the species to a group of chemicals with the same MOA. Single and multiple linear regressions between the MSS values and species trait data at the family level explained up to 70% of the variation in invertebrate sensitivity to 3 groups of insecticides. Recently, the approach by Rubach, Baird and Van den Brink (2010) has been extended by incorporating relationships between species traits and sensitivity into predictive models that could potentially be used in risk assessment (Rico and Van den Brink 2015).

So far, these predictive models have only been built for insecticides, whilst it remains open whether the same traits are important in explaining invertebrate sensitivity to other groups of chemicals. Therefore, in this study, we develop predictive models of macroinvertebrate sensitivity to a larger and more diverse set of MOAs. We first optimize the sensitivity prediction method of Rubach et al. (2010) and Rico and Van den Brink (2015) using both literature research and comparative analysis. The methods, results and discussion of this technical optimization are given in the Supplementary Information (S1). Next, and as the main focus of this article, we study the influence of MOA on species sensitivity rankings and, subsequently, on the resulting predictive trait models. By studying which traits are included in the trait-sensitivity models, we aim to get a better mechanistic understanding of how species sensitivity is determined by MOA. In a final evaluation, we test the power of species sensitivity predictions and assess if model performance depends on data availability. As a result of our research, we deliver a trait-based macroinvertebrate sensitivity modelling tool that builds predictive models for potential use in risk assessment for a diverse set of pre-defined MOAs.

2.2 Methods

2.2.1 Overall methodology

The prediction tool was developed in the R environment, and automates sensitivity rankings of aquatic macroinvertebrates and the construction of a corresponding set of trait-based models from a MOA-based chemical grouping (**Figure 2.1**). A hands-on explanation of the R tool and all R scripts and databases required to run the tool are available in the Supporting Information (S1 and S2 respectively). The tool consists of four different parts: database collection, preliminary input data processing, data processing, and production of output. Each part in turn, consists of multiple steps, the essentials of which will be presented briefly.

2.2.2 Database collection and preliminary input data processing

Several public databases were utilized to obtain data on MOA, toxicity, chemical properties, traits and taxonomy. We accept the limitations and errors of the databases exploited in this study, as it is not within the scope of this study to perform a quality assessment on the databases used.

The extensive (1213 chemicals) MOA database developed by Barron et al. (2015) was used to extract a list of chemicals with their CAS numbers and their division into 6 groups of broad MOA (narcosis, acetylcholinesterase (AChE) inhibition, ion/osmoregulatory/circulatory (IOC) impairment, neurotoxicity, reactivity, and electron transport inhibition), and their subdivision into 31 groups of specific MOA (see Table 1 in Barron, Lilavois et al. 2015). The database was cleaned (e.g. chemicals for which no specific MOA was defined were removed) and prepared for further processing (e.g. specific MOAs with long names were replaced by a letter to ensure readability). For more details on pre-processing, see the corresponding R script in the SI (S2).

The US Environmental Protection Agency (EPA) ECOTOX database (USEPA 2017) was selected as the source of toxicity data. See instructions in S1 on how to download the ECOTOX database. Only the ECOTOX tables *tests*, *results* and *species* are incorporated into the R tool.

A chemical properties database was acquired by batch-running all CAS numbers available in the MOA database in the EPI (Estimation Programs Interface) Suite programs KOWWIN, MPBPVP and WSKOWWIN, respectively obtaining data on logKow (octanol-water partitioning coefficient); melting point, boiling point and vapor pressure; and water solubility (USEPA 2018). Modelled values were only used when experimental data were lacking. Eventually, only data on water solubility are used in the tool (as a check for realistic concentration values). All data are kept in however, to enable future flexibility of the tool. Data on molecular weight were obtained and added to the chemical properties database by extracting SMILES (Simplified Molecular-Input Line-Entry System) of all MOA CAS numbers from the SMILECAS database (also available through EPI Suite), and calculating the molecular weight based on these SMILES using the **rcdk** package in R (version 3.4.5, Guha 2007).

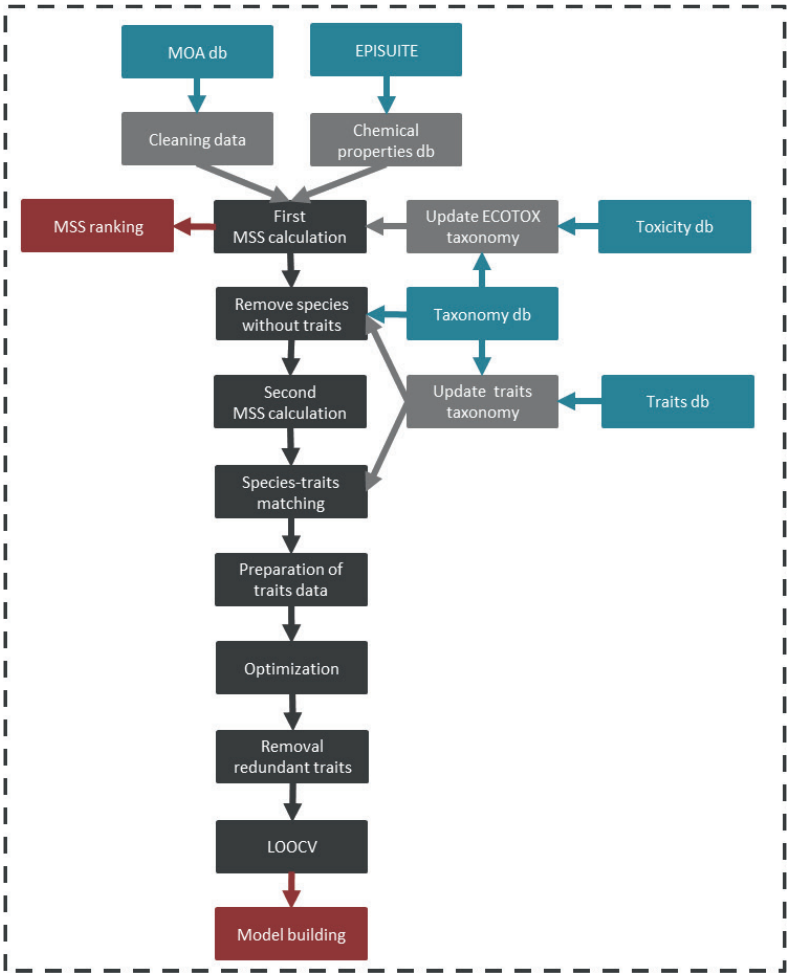


Figure 2.1. Structure of the developed R tool, divided into the four different parts: database collection (blue), preliminary input data processing (light grey), data processing (dark grey) and production of output (red). MSS refers to Mode Specific Sensitivity; LOOCV refers to Leave-One-Out-Cross-Validation.

The Tachet database provides a coding of 22 biological and ecological traits describing 472 species known to live in French freshwaters (Tachet, Richoux et al. 2000, Usseglio-Polatera, Bournaud et al. 2000). The database is based on '*a very large and scattered published expert knowledge and diverse literature sources ... We also included unpublished observations of ourselves and colleagues*' (Usseglio-Polatera, Bournaud et al. 2000). From this, we deduced that the database represents the average trait state of European species and is, therefore, suitable for our study. Since our interest lies in predicting species sensitivity and we want to avoid overfitting, only the following traits for which we could hypothesize a mechanistic relation with sensitivity were extracted (based on Rubach, Baird et al. 2010): maximum potential size (i.e. ≤

0.25 cm, > 0.25 – 0.5 cm), life cycle duration (i.e. ≤ 1 year, > 1 year), potential number of cycles per year (i.e. < 1, > 1), dispersal mode (i.e. aquatic passive, aerial active), respiration mode (i.e. tegument, gill), feeding mode (i.e. shredder, scraper), current velocity preferendum (i.e. slow, fast), salinity preferendum (i.e. freshwater, brackish water), temperature preferendum (i.e. < 15°C, > 15°C) and pH preferendum (i.e. ≤ 4, > 5.5-6) (see Table S1 for an overview of the traits and the corresponding trait categories).

To facilitate the cross-linking of information among the different databases, the Taxonomy database of the NCBI (National Centre for Biotechnology Information, Benson, Karsch-Mizrachi et al. 2009, Sayers, Barrett et al. 2009) was used to extract the scientific names, along with the taxonomic rank and unique id, of all the species present in both the ECOTOX and the Tachet database. For this we used the **taxize** package in R (version 0.9.0, Chamberlain and Szöcs 2013). A copy of the ECOTOX *species* table and the Tachet database, both with updated taxonomy, are provided in the SI (S2).

2.2.3 Data processing

Data processing has been executed once for each broad and once for each specific MOA (36 times in total). The data processing part of the tool consists out of eight consecutive steps (**Figure 2.1**): i) first Mode Specific Sensitivity (MSS) calculation, ii) removal of species without traits information, iii) second MSS calculation, iv) species-traits matching, v) preparation of trait data, vi) optimization, vii) removal of redundant traits, and, viii) Leave-One-Out-Cross-Validation (LOOCV).

The MSS values are calculated twice; once including all species for which sufficient toxicity data are available (step i), and once only including species for which we have ascertained that also trait data are available (step iii). In between the two MSS calculations, species for which no trait data are available are removed (step ii). Performing the MSS calculation twice is necessary, because the MSS values depend on relative sensitivities, and therefore are influenced by the in- or exclusion of species. The MSS calculations are implemented as described by Rubach, Baird and Van den Brink (2010). In short, only toxicity tests lasting between 1 and 4 days and studying the effect of chemical stress on mortality are included in the analysis. Within each chemical, first the log transformed LC50 or EC50 values are normalized using the mean and standard deviation of all sensitivity values found for that chemical, resulting in a relative sensitivity of each species towards that chemical. Subsequently, these relative sensitivities are averaged over all chemicals belonging to the same MOA, resulting in an MSS value. Importantly, there were three fundamental differences in our method compared to the method of Rubach, Baird and Van den Brink (2010): the unit of the toxicity data (we use mol/L instead of µg/L), an extra check for realism of concentration values (should not exceed solubility), and the selection of tests with the longest exposure duration to enlarge the chance of reaching equilibrium concentration (see discussion for more extensive explanation).

After the second MSS calculation, species-traits matching is executed at the lowest taxonomic level possible (step iv). Since species level trait data are scarce, and traits measured at genus and species level are strongly correlated (Dolédec, Olivier et al. 2000), species-traits matching is done at genus level. Prior to species-traits matching, MSS values are converted to genus level by averaging the MSS values of all species belonging to the same genus. Traits data are also converted to genus level by taking the median of the original fuzzy codes of all taxa belonging to that genus. This fuzzy coding scheme is specifically developed for describing species traits data, and is ideal for differentiating species by their affinities to different trait modalities (i.e. categories) belonging to several traits (for further explanation of fuzzy codes, see Chevenet, Dolédec et al. 1994).

Next, the trait data are prepared for linear regression by expressing continuous traits (e.g. size) as weighted averages of the different trait modalities (Tachet, Richoux et al. 2000, Rubach, Baird et al. 2010, Rico and Van den Brink 2015), and factorial traits (e.g. mode of respiration) as fixed within species (step v) (Rubach, Baird et al. 2010, Rico and Van den Brink 2015). The latter means that for each of the factorial traits, the modality for which the taxon has the highest affinity is selected as the modality this taxon carries and uses throughout its entire lifespan. In case of equal affinity to more than one trait modality (e.g. 40% gills, 40% skin respiration), a missing value is inserted. These missing values are problematic for (multiple) linear regression, because regression requires closed datasets with no missing values. To solve this, an optimization step is performed (step vi), which removes all gaps from the dataset by deleting any species with missing trait values.

Next, redundant traits are removed to avoid overfitting (step vii). Traits are considered redundant i) when they are clearly aliased with other traits (collinearity), and ii) when they do not show enough variation in their different trait modalities. The tool tackles these two issues through i) a collinearity maximum of 0.7 (see Dormann, Elith et al. 2013 for a review of different methods to deal with collinearity), removing all traits that exceed this maximum, starting with those that correlate with the largest number of other traits or, in case multiple traits match this criterion, with the largest exceedance of the collinearity maximum, and through ii) a minimum on the trait modality diversity index (derived from the Shannon diversity index, Spellerberg and Fedor 2003), removing all traits with a trait modality diversity below this minimum.

Finally, a Leave-One-Out-Cross-Validation (LOOCV) is executed to quantify the predictive power of the model (step viii). LOOCV is done by successively leaving out one species from the training dataset and building the MSS model based on the remaining species (for explanation of model building, see section 2.2.4). LOOCV was preferred above k-fold cross-validation, because LOOCV is better in giving a reliable picture of the true R^2 than k-fold cross-validation (Hawkins, Basak et al. 2003). With the produced model, the MSS value of the left-out species is predicted, and afterwards compared to the known value by calculating the squared

error. From the LOOCV, the mean squared prediction error (MSPE) and a prediction coefficient (P^2) were calculated as follows:

$$MPSE = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n} \quad \text{eqn. 1}$$

$$P^2 = 1 - \frac{MSPE}{s_y^2} \quad \text{eqn. 2}$$

where $\hat{y}_i$ is the estimated MSS value, y_i is the calculated MSS value, and s_y^2 is the variance of all MSS values in the dataset. When all predictions perfectly match observations, P^2 equals 1. A negative P^2 value indicates that prediction errors exceed the total variance of the sensitivity values, and therefore the model has poor accuracy.

2.2.4 Production of output

Each MOA output consists of the MSS ranking and the best MSS model. The MSS ranking is obtained by sorting the MSS values found in the first MSS calculation from low (most sensitive) to high (least sensitive). To visualize the taxonomic patterns in species sensitivity, we made heat maps of the MSS rankings, dividing the MSS values into four bins, ranging from sensitive to tolerant (MSS ≤ -1; -1 – 0; 0 – 1; ≥ 1). The best MSS model is selected in two steps. First, all possible linear models are constructed using the **regsubset** function of the R package **leaps** (version 3.0, Lumley and Miller 2017). Next, the best model is selected from all constructed models based on the small sample unbiased Akaike's Information Criterion (AICc), which takes both model fit and model complexity into consideration and which is additionally extended with a bias correction term for small sample size (Johnson and Omland 2004).

2.3 Results and discussion

Figures of all MSS rankings resulting from the first MSS calculation and tables comparing within and among broad MOA relative sensitivities can be found in S1 and S3 respectively. For the interpretation of the results, it is important to realize that the MSS values represent the relative sensitivity of a species to a group of chemicals with the same MOA. The results of the MSS rankings (section 2.3.1) and the MSS models (section 2.3.2) are not described together, because the MSS models result from a further reduced and transformed MSS dataset (**Figure 2.1**).

2.3.1 MSS rankings

We found large differences in the MSS rankings of species depending on MOA, both within the broad MOAs (between the specific MOAs belonging to one broad MOA), and between the different broad MOAs. Differences between broad MOAs were, however, larger than differences within broad MOAs. This can be seen when comparing the standard deviation (SD) of the MSS values across the broad MOAs (0.68), with the SD of the MSS values within the broad MOAs (0.26, 0.38, 0.4, 0.52, 0.59, and 0.63 for within respectively the broad MOAs

AChE inhibition, narcosis, IOC impairment, electron transport inhibition, neurotoxicity, and reactivity) (S3). Having a higher SD between the broad MOAs compared to within the broad MOAs confirms that grouping into specific MOAs is helpful for modelling species sensitivity (Barron, Lilavois et al. 2015, Martin, Young et al. 2015, Kienzler, Barron et al. 2017).

A heat map of the MSS rankings shows the general taxonomic pattern of species sensitivity and tolerance to the different MOAs (**Figure 2.2**). For AChE inhibition, for instance, arthropods are most sensitive and molluscs are most tolerant. Whether crustaceans or insects are the most sensitive arthropods to AChE inhibition remains unclear, since both groups contain comparable numbers of sensitive and tolerant genera. We see similar results in two closely related studies (Rubach, Baird et al. 2010, Rico and Van den Brink 2015), as well as in a review of semi-field experiments (Wijngaarden, Brock et al. 2005). The sensitivity pattern for narcotic chemicals is closely related to the sensitivity pattern of AChE inhibition, with crustacean and insect genera containing the largest number of sensitive genera. We compared this result with two outdoor microcosm studies performed with the fungicide azoxystrobin, classified as a narcotic chemical (Barron, Lilavois et al. 2015). In the study of Zafar et al., (2012) significant effects were found for only one insect species (*Chaoborus obscuripes*) and for none of the crustacean species tested. In the study from Cole et al. (2000) (summarized in EFSA 2009), the mollusk Sphaeriidae was the only species showing negative effects on occurrence after a constant exposure of 10 µg/L, followed by negative effects on Gammaridae, Oligochaeta and Planorbidae at 30 µg/L, and on Asellidae at 100 µg/L. No negative effects on any of the insect groups tested (Hemiptera, Chaoboridae, Chironomidae) were found. The discrepancy between our results and the two microcosm studies might be due to the wrong assignment of narcosis as the MOA of azoxystrobin. Indeed, although two studies (Bradbury 1994, Barron, Lilavois et al. 2015) classified azoxystrobin as a narcotic chemical, two other classification schemes (the QSAR Toolbox developed by the OECD, and Toxtree) assigned a non-narcotic MOA to azoxystrobin (Kienzler, Barron et al. 2017). In the last paragraph of section 2.3.3 we discuss the causes and effects of MOA misclassification in more detail.

Besides general patterns in species sensitivity, the heat map also shows that there is not one genus, family or class which is sensitive to all MOAs (**Figure 2.2a**). When looking over all MOAs, the arthropod phylum contained the largest fraction of species classified as being more sensitive than average, followed by nematodes, molluscs and annelids. Genera belonging to the phyla Bryozoa, Cnidaria, Platyhelminthes and Rotifera were never more sensitive than average. At class level, Hexanauplia contained the largest fraction of genera more sensitive than average, which makes sense, because of their relatively small size and, herewith, large surface to volume ratio. Another result is that common test taxa are often not the most sensitive at all (**Figure 2.2b**). We find that Daphnia never belongs to the most sensitive group, and other commonly tested taxa (e.g. Asellus, Procambarus, and Chironomus) are for the majority of the MOAs found to be medium tolerant. Only the commonly tested taxon Ceriodaphnia shows high sensitivity to two MOAs: AChE inhibition and reactivity.

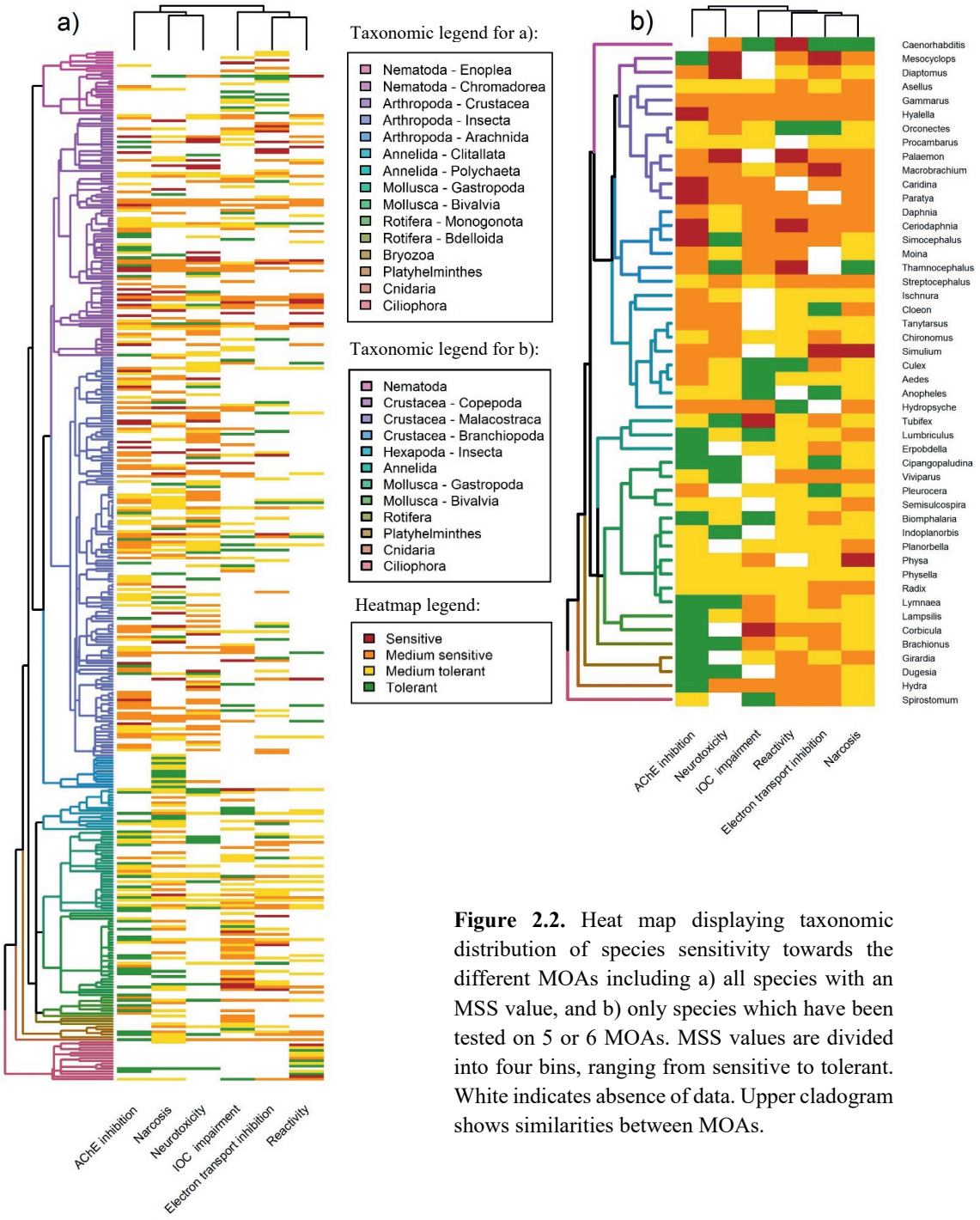


Figure 2.2. Heat map displaying taxonomic distribution of species sensitivity towards the different MOAs including a) all species with an MSS value, and b) only species which have been tested on 5 or 6 MOAs. MSS values are divided into four bins, ranging from sensitive to tolerant. White indicates absence of data. Upper cladogram shows similarities between MOAs.

The information presented in **Figure 2.2** can also be used to identify for which MOA-taxon combinations data is lacking. Ciliophora have, for instance, been studied more frequently for reactive chemicals, whilst Nematodes have been studied more frequently for the MOAs IOC impairment and electron transport inhibition. Generally, the MOAs IOC impairment, electron transport inhibition and reactivity have been studied to a lesser extent than the others. Especially data on insect species are missing for these MOAs, which might explain the counterintuitive result of seemingly reduced sensitivity of insects to these MOAs. Indeed, a field study performed in the 1970s shows that two mayfly (Ephemeroptera) species significantly reduced in occurrence after application of the electron transport inhibitor Antimycin A, whilst all other invertebrate taxonomic groups present at the study site remained unaffected (Morrison 1979). This indicates that certain insect species are indeed sensitive to electron transport inhibition, but these species are not included in the ECOTOX database. That the taxonomic focus of some MOAs differs can lead, besides to a misinterpretation of results as just has been demonstrated, to a bias in the sensitivity ranking, and subsequently, to a bias in the sensitivity models. That the taxonomic composition of the species assemblage used to construct models is important is well known. Maltby et al. (2005) found, for example, that hazardous concentrations (HC5) derived from arthropod SSDs were significantly lower than those derived from non-arthropod invertebrates. In order to avoid any bias in predictive modelling, we should, therefore, not only ensure high taxonomic coverage, but also evenness across the different taxonomic groups.

2.3.2 Sensitivity models

The final modelling effort resulted in 12 significant ($p < 0.05$) models explaining 31 to 90% of the variation in MSS values, covering 5 broad MOAs and 7 specific MOAs (**Table 2.1**). Some of the MOAs show overlap in the traits that explained sensitivity best. For example, life cycle duration (life) and feeding mode (feeding) are included in the models for 3 of the 6 broad MOAs. This is a good indicator that these traits are in general important in explaining species sensitivity. Results from several studies (Rubach, Baird et al. 2010, Ippolito, Todeschini et al. 2012, Rico and Van den Brink 2015) confirm this by including the same or closely related traits in their models. Additionally, an extensive review of potential sensitivity related traits classify feeding mode and life cycle duration as traits known to have an established link with sensitivity for several taxa (Rubach, Ashauer et al. 2011). Other traits selected by our modelling effect (e.g. temperature and salinity preferendum) are only included in explaining one broad MOA, which might indicate that these traits are less important for determining species sensitivity. For temperature preferendum, this is indeed confirmed by the relatively low prediction coefficient of the models including this trait (**Table 2.1**). One of the models containing salinity preference (alicyclic GABA antagonism), however, has one of the highest prediction coefficients, and is therefore possibly important. Indeed, a relationship between salinity preference and the toxicity of a GABA antagonist makes sense, because GABA is one of the most common neurotransmitters (Nelson and Cox 2008), and is therefore potentially influenced by the presence of strategies to deal with salt stress (Rivera-Ingraham and Lignot 2017).

Table 2.1. Model coefficients of the best models (smallest AICc) that were found significant ($p \leq 0.05$) for the different MOAs using exhaustive linear regression analysis. Model fit is shown as the adjusted R^2 (R^2), and predictive power is shown as the prediction coefficient (P^2). See table S1 for an explanation of the traits and trait modalities used in this analysis.

Broad MOA	Specific MOA	pH pref.	Dispersal mode	Respiration mode	Life cycle dur.	Life cycles yr^{-1}	Feeding mode	Temp. pref.	Max. potential size	Velocity pref.	Salinity pref.	R^2	P^2
Narcosis		-0.32	-0.42	0.51								0.33	0.013
	Nonpolar				-0.77							0.42	-0.711
	Polar		-0.44	0.53								0.36	0.189
Neurotoxicity				-0.35		1.26	0.16				2.44	0.31	-0.29
	Alicyclic GABA antagonism										6.13	0.44	0.193
AChE inhibition					-1.11	-0.75	-0.17	-0.92				0.41	-0.027
	Organophosphate				-0.66		-0.15	-0.83				0.33	-0.33
	Carbamate	1.16			-0.74	-1.48						0.62	-0.125
Reactivity			0.46									0.67	0.255
	Chromate								-0.91	-1.26		0.9	-1.611
ETI[a]	Uncoupling oxidative phosphorylation						0.25		-0.64			0.41	-0.446
IOC impairment						-1.6				1.04		0.48	0.326

[a)] Electron Transport Inhibition

For some traits, the correlation to sensitivity is positive for some MOAs and negative for other MOAs (e.g. dispersal, **Table 2.1**). Species were more sensitive to narcosis when they preferred more neutral waters (pH preferendum), were capable to disperse actively (dispersal) and breathed through their tegument (respiration mode, see Table S1 for the included trait modalities). For reactivity, however, we find that the relationship between dispersal and sensitivity is the exact opposite, and the more passive organisms were found to be more sensitive. This can be explained by the large gap in insect genera for the MOA reactivity (**Figure 2.2**). When comparing the data included in the multiple linear regression, the dataset for reactivity only contains some insensitive Odonata and Diptera genera that are classified to disperse actively, whilst the dataset for narcosis contains both sensitive Ephemeroptera and sensitive Trichoptera genera. Because of this unbalance, it remains unclear whether the direction of the trait-sensitivity relationship of the two MOAs is indeed mechanistic, or it is simply an artefact of data availability. We think that the latter is true, because the trait profile of species sensitive to narcotic chemicals matches perfectly with the trait profiles of insects belonging to the orders Ephemeroptera, Plecoptera and Trichoptera (EPT), which are generally known as sensitive species (Verberk, Van Noordwijk et al. 2013). It can be hypothesized that our trait selection does not contain the best traits, and the patterns we see only arise because traits are phylogenetic correlated, i.e. belong to trait syndromes (Poff, Olden et al. 2006, Verberk, Van Noordwijk et al. 2013). Further analyses including phylogenetic information can probably further elucidate this.

2.3.3 Data gaps and potential for improvement

Low availability of traits data has in earlier studies been described as one of the major threats to the successful implementation of trait-based approaches in ecotoxicology (Rubach, Ashauer et al. 2011, Van den Brink, Alexander et al. 2011). We confirm this by showing that the lack of trait data was the main obstacle in model construction. Over all the MOAs, an average of only 12% (SD ±5%) of the species for which we had sufficient data to calculate the first MSS value were included in the construction of the MSS models. For an average of 56% (±10%) of the species, no match at all could be found in the trait database. For 15% (±5%) of the species, a match could be found in the trait database, but values on one or multiple traits of interest were lacking, and the species were therefore removed during the optimization process. It is generally known that traits databases hold a high number of gaps, and 15% of the species with missing values matches closely to 18% of the species having an incomplete trait description in a comparable trait database for Europe (Gayraud, Statzner et al. 2003). Due to the large loss of taxa in the analysis because of missing traits information, also only 20% (±7%) of the toxicity tests used to calculate the first MSS value ended up as an underlying data point in the multiple linear regression analysis. This indicates that a large part of the available toxicity data remains unused, merely because of insufficient trait data. Future research should therefore focus on completing, and simultaneously extending the taxonomic extent of existing trait databases. This can be achieved by obtaining new data, but also by combining existing trait databases, although

the latter could become complicated due to existing differences in methodologies and categories between the different trait databases. Baird et al. (2011) review the technological challenges in creating and sharing traits data; and Culp et al. (2011) show examples of traits databases currently available for different habitats and taxonomic groups (cf. Table 2).

In our analysis 16% of the traits we deemed important regarding species sensitivity did not enter the multiple linear regression process, because they were highly correlated to another trait. This could indicate the permanent presence of collinearities between some of the sensitivity-related traits we selected (also defined as trait-syndromes, Poff, Olden et al. 2006). However, there was a clear relationship between the number of traits and the number of species going into the multiple linear regression process, indicating that collinearity was primarily found in small datasets (Figure S4 in the Supporting Information of Chapter 2). Therefore, traits were in most cases only correlating with each other due to chance, originating from the small size of the remaining datasets after all data processing steps. Besides the sensitivity-related traits selected and evaluated in this study, it can be helpful to enhance trait databases with finer morphological characteristics which are more directly related to bioaccumulation (e.g. lipid content), or to the internal distribution of the chemicals (e.g. target site distribution). Buchwalter and Luoma (2005) found, for example, that although well-studied traits like body size or gill size did not explain macroinvertebrate sensitivity to heavy metals, the relative number of ionoregulatory cells did relate to dissolved metal uptake rates. Rubach et al. (2012) also tried to improve trait-models by using finer morphological traits. Additionally, they measured most of these traits on individuals captured in the same season (although a different year) and in the same biogeographic region (sometimes even the exact same location) as they were collected for toxicity testing, which reduces the presence of trait variability due to phenotypic or seasonal intraspecific variability. However, their results show only a minor increase in model fit by measuring traits, even with the lowered intraspecific variability, both for the single linear regression and for the multiple linear regression (compare Table 2 in Rubach, Baird et al. 2010, with Table 3 and Figure 2 in Rubach, Baird et al. 2012).

There are other aspects that could improve trait-based models, three of which we will discuss in more detail. First, measurement of internal tissue concentrations (i.e. the tissue residue) in addition to external exposure concentrations would improve toxicity predictions, because toxic effects are more closely related to the concentration inside the organism than the concentration in the water (Friant and Henry 1985, McCarty, Landrum et al. 2011). Although this has already been proven in the eighties, research on the (partial) incorporation of tissue residues into ERA has only started the previous decade (e.g. in the biotic ligand model, Paquin, Gorsuch et al. 2002), and has still not been accepted as a standard in chemical regulation. We hope that the recent development and increasing importance of mechanistic effect models like toxicokinetic (TK) – toxicodynamic (TD) models (e.g. the General Unified Threshold Model of Survival, Jager, Albert et al. 2011, or the Dynamic Energy Budget, Jager, Martin et al. 2013), will lead to a more frequent measurement and publication of internal concentrations. Especially since

these process-based models explain the link between internal concentrations and dynamic processes underlying toxic responses, and additionally enable extrapolation from standard test conditions, e.g. to different exposure scenarios (Ashauer, Boxall et al. 2007, Ashauer, Boxall et al. 2007, Ashauer, Hintermeister et al. 2010). Rubach et al. (2012) additionally demonstrated that quantitative links between the parameters of these mechanistic effect models fitted on internal concentration data and traits are substantially stronger than quantitative links between classical sensitivity endpoints (e.g. EC50, LC50) and traits. Measurements of internal concentrations could therefore improve our models in two ways: one, by using internal instead of water concentrations, and two, by trying to explain TKTD parameters, instead of LC50 values, which could subsequently be used to model EC50 values at any exposure profile of interest.

Second, and especially important in the absence of internal concentrations, the exposure duration of toxicity tests should be long enough to ensure approximate equilibrium between external and internal concentrations. Without going too deep into the discussion of acute versus chronic exposure, biotransformation, and elimination, we do want to emphasize that failure to obtain equilibrium concentration can result in an underestimation of effects, simply because under the same exposure time, a smaller organism reaches a higher internal concentration faster than a larger organism (McCarty, Landrum et al. 2011). This means that for an acute toxicity test, different organisms may require a different minimum exposure time. We tried to account for this by selecting the toxicity test with longest exposure duration where possible. However, it can still be that some of the variability we see in **Figure 2.2** may be associated with differences in exposure duration (e.g. in one toxicity test, a species was exposed for 48 hours to chemical A, whilst in another toxicity test, the same species was exposed for 96 hours to chemical B).

Finally, improvements can be made by integrating Bayesian methods into the current approach. Approximate Bayesian Computation can distinguish which mechanisms contribute most to patterns observed in the data (Hartig, Calabrese et al. 2011), and can thereby help optimizing model complexity. Bayesian model averaging can additionally incorporate model-selection uncertainty into statistically derived predictive models, and can therefore provide a more realistic estimation of model uncertainty compared to the approach used in our study (Wintle, McCarthy et al. 2003, Clark 2005).

Interestingly, data availability alone could not explain differences in model performance. An increase in the number of toxicity tests, chemicals or genera did not improve model performance substantially (Figures S1, S2, and S3 in the Supporting Information of Chapter 2). Actually, model fit reduces slightly with an increase in the number of tests, chemicals or genera. However, cross-validation results show an exact opposite trend, with a slight reduction in error with an increase in data availability. We think that model performance does not increase with data availability due to one (or a combination) of the following three reasons. First, the chemical

groups (in our case, the MOAs) may insufficiently differentiate the chemicals according to the effects that they cause in invertebrates. As in any kind of grouping, mistakes or missing information could result in a wrong grouping in MOA for the chemicals (see Martin, Young et al. 2015, Kienzler, Barron et al. 2017 for studies on MOA classification errors). Additionally, MOA is not a constant property of a compound, but may vary between species or life stage, depending, for instance, on the availability of target sites, exposure duration or frequency, or the endpoint of interest (Nendza and Muller 2000). Photosynthetic inhibitors, for example, are specific toxicants towards primary producers, but are often baseline toxicants towards invertebrates (Nendza and Muller 2000). The chance that a chemical expresses multiple MOAs in different species becomes smaller, however, when analysis is restricted to species belonging to the same organism group (e.g. invertebrates, as in this study), because the basic biochemical systems and molecular targets affected by each MOA may be generally conserved across many species (LaLone, Villeneuve et al. 2016). A second aspect that could have prevented increased model performance with increased data availability can be that we missed the right traits to mechanistically explain the relationship between MOA and sensitivity. As mentioned before, obtaining additional and finer morphological traits can help improve trait-based models, especially if these traits explain the toxicodynamics of the chemical, since processes related to toxicodynamics are currently not covered by the traits in our selection. Finally, we argue that the models could not achieve a better model performance because, regardless of data availability, trait data alone may be insufficient in explaining species sensitivity. Several studies show that complementing trait data with data on phylogenetics greatly enhances model performance, and that both traits and phylogenetic indicators explain a distinct part of species sensitivity (Poteat, Jacobus et al. 2015, Pilière, Verberk et al. 2016). Other attempts combining phylogenetics and physiochemical properties into predictive models have also proven successful (Guénard, von der Ohe et al. 2014, Malaj, Guénard et al. 2016), although phylogenetics is still impossible to include in an inclusive approach as we performed in this study. Nevertheless, we suggest to further explore the potential of species traits for sensitivity predictions, as they enable large scale applicability and increased mechanistic understanding.

2.3.4 Future direction

Our prediction tool is an addition to the fast growing and evolving science behind environmental risk assessment, which can be easily amended in the future when more data or better data processing and analysis procedures are identified. It is very flexible, and facilitates i) testing the effectiveness of different kinds of chemical grouping, e.g. based on (physical) mechanism of action (Ashauer and Jager 2018), ii) testing the predictive value of ‘new’ traits, hypothesized to have a relationship with species sensitivity, iii) the addition of new predictors, e.g. based on phylogenetics or physiochemical properties, and iv) repeating the modelling exercise for different groups of organisms (e.g. fish or algae), as long as trait data are available. Analyses comparable to those reported herein would take days to weeks of preparation time to collect all input data, followed by more time to conduct and compile all analyses, requiring

multiple types of commercial software. Using the software and algorithms developed for this work, these intensive efforts can now be compiled in only a few hours, already has all main open-source databases incorporated and uses only R, a free software environment.

Both the sensitivity rankings and the sensitivity models produced in this study are an important step forward on the challenging road towards predictive ecotoxicology with reduced animal testing (Villeneuve and Garcia-Reyero 2011). The sensitivity rankings help reduce animal testing by guiding future taxonomic focus of toxicity tests depending on the estimated MOA of the new chemicals. The sensitivity models can predict the sensitivity of species or even entire communities never tested before, avoiding additional animal tests. This allows us to rank the sensitivity of species occurring in any community composition anywhere, or in other words, to determine the worst-case ecological scenario for any aquatic ecosystem. Developing such ecological scenarios will fill one of the remaining gaps in the construction of environmental scenarios (De Laender, Morselli et al. 2015, Franco, Price et al. 2016, Rico, Van den Brink et al. 2016) deemed necessary for future ecological risk assessment frameworks. Together with exposure scenarios, ecological scenarios will form the basis for developing spatial-temporal explicit population-, community- and ecosystem-level effect models for use in prospective environmental risk assessment for chemicals.

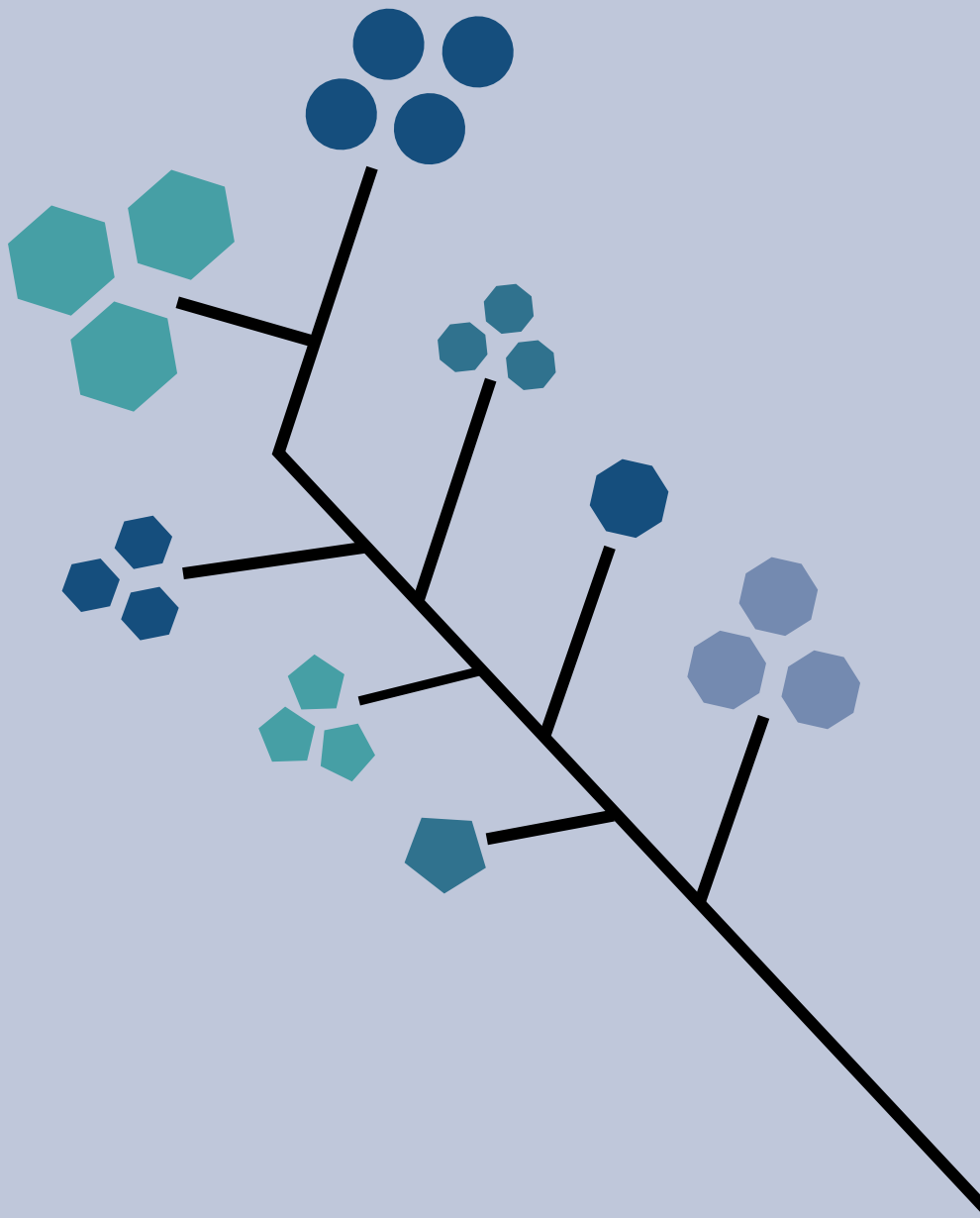
Acknowledgements

This project has been funded by Unilever under the title 'Ecological Scenarios and Models for Use in chemical risk assessment' (ESMU). Thanks to Pablo Rodriguez for assisting in programming issues, to Merel Kooi with help in ggplot, and to Lara Schuijt for commenting on schematic figures. The authors would also like to express their gratitude to the editors involved and the anonymous reviewers for their insightful comments and suggestions.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:10.1021/acs.est.9b00893.

(S1) PDF file: i) Hands-on explanation of the R tool, ii) Downloading instructions for the ECOTOX database, iii) Table S1. Description of traits and trait categories, iv) Figures S1 – S4. Relationships between model input and model performance, v) Comparison of modelling methods and decisions, vi) Figures S12 – S42. MSS ranking for each Broad and Specific MOA. (S2) Zip folder: i) All required R scripts, ii) Copy of required databases. (S3) Excel file: i) Tables comparing within and among broad MOA rankings, ii) Modelling sets resulting from running the R tool using the different practices.



CHAPTER 3

Potential impact of chemical stress on freshwater invertebrates: A sensitivity assessment on continental and national scale based on distribution patterns, biological traits, and lineage.

Sanne J. P. Van den Berg, Cecilie Rendal, Andreas Focks, Emma Butler, Edwin THM Peeters, Frederik De Laender, Paul J. Van den Brink

Abstract

Current chemical risk assessment approaches rely on a standard suite of test species to assess toxicity to environmental species. Assessment factors are used to extrapolate from single species to communities and ecosystem effects. This approach is pragmatic, but lacks resolution in biological and environmental parameters. Novel modelling approaches can help improve the biological resolution of assessments by using mechanistic information to identify priority species and priority regions that are potentially most impacted by chemical stressors. In this study we developed predictive sensitivity models by combining species-specific information on acute chemical sensitivity (LC50 and EC50), traits, and taxonomic relatedness. These models were applied at two spatial scales to reveal spatial differences in the sensitivity of species assemblages towards two chemical modes of action (MOA): narcosis and acetylcholinesterase (AChE) inhibition. We found that on a relative scale, 46% and 33% of European species were ranked as more sensitive towards narcosis and AChE inhibition, respectively. These more sensitive species were distributed with higher occurrences in the south and north-eastern regions, reflecting known continental patterns of endemic macroinvertebrate biodiversity. We found contradicting sensitivity patterns depending on the MOA for UK scenarios, with more species displaying relative sensitivity to narcotic MOA in north and north-western regions, and more species with relative sensitivity to AChE inhibition MOA in south and south-western regions. Overall, we identified hotspots of species sensitive to chemical stressors at two spatial scales, and discuss data gaps and crucial technological advances required for the successful application of the proposed methodology to invertebrate scenarios, which remain underrepresented in global conservation priorities.

This chapter is based on the paper: Sanne J. P. van den Berg, Cecilie Rendal, Andreas Focks, Emma Butler, Edwin T.H.M Peeters, Frederik De Laender, Paul J. Van den Brink (2020). Potential impact of chemical stress on freshwater invertebrates: A sensitivity assessment on continental and national scale based on distribution patterns, biological traits, and relatedness. Science of the Total Environment, Accepted.

3.1 Introduction

The scientific community is rapidly developing new ecological models to increase realism in environmental risk assessment (ERA, e.g. De Laender, Morselli et al. 2015, Windsor, Ormerod et al. 2018). However, what so far has remained unclear is which organisms need to be modelled. Common standard test species are usually not representative of all species present in ecosystems with regards to their sensitivity to stressors (Nagai 2016). Indeed, it has already been argued for over 30 years that there is not a single species or a specific group of species which is always the most sensitive (all the time, everywhere, and towards every compound). This has been coined the 'myth of the most sensitive species' (Cairns 1986). However, since in reality both compound multiplicity as well as species diversity occur simultaneously, it is not feasible to acquire all possible sensitivity data with laboratory toxicity testing. Therefore, there is a need to develop models that can help identify priority species, which are species that are likely to be intrinsically most sensitive to chemical stressors.

Several studies have tried to determine which species are intrinsically most sensitive to chemical stressors by using species traits, and were able to explain up to 87 percent of the variation in species sensitivity using only four traits (Rubach, Baird et al. 2010, Rubach, Baird et al. 2012, Rico and Van den Brink 2015, Van den Berg, Baveco et al. 2019). A large advantage of using traits-based approaches is that they add mechanistic understanding of the sensitivity process by describing characteristics that make a species more or less sensitive towards chemical stressors. This largely reduces the chances of overfitting models to the training data (Johnson and Omland 2004). In addition to that, describing aquatic communities in terms of their biological traits increases the generality of such characterizations and their subsequent transferability between regions (Van den Brink, Alexander et al. 2011). Also, correlations between species traits and species sensitivity might exist, potentially resulting in unexpected effects at the community level (Baert, De Laender et al. 2017).

Other studies (Malaj, Guénard et al. 2016) concerned with determining which species were most sensitive to chemical stressors, combined phylogenetic information with chemical properties. They were to a great extent (R^2 of ~0.8) capable of predicting species sensitivity to pesticides (Guénard, von der Ohe et al. 2014) and heavy metals (Malaj, Guénard et al. 2016). Furthermore, some studies have demonstrated that indeed traits and phylogeny (or other measures of relatedness between species) both explain an unique part of the sensitivity process (Poteat, Jacobus et al. 2015, Pilière, Verberk et al. 2016). However, phylogenetic approaches do not unravel any concrete mechanisms of sensitivity, and are therefore more susceptible to overfitting on the training data. For this reason, we think that a combination of both traits and phylogenetic information has the most potential for identifying priority species at a large spatial scale.

We envision these priority species to, in the future, become part of environmental scenarios, a simplified (model) representation of exposed aquatic ecosystems which provides a sufficient amount of ecological realism, enabling us to conduct an appropriate ERA (Rico, Van den Brink et al. 2016). There are clear benefits associated with the development of scenarios for use in risk assessment, the most important ones being reduction of animal tests, integration of exposure and effect assessments, and increased realism with respect to spatial-temporal dimensions and species biodiversity (Rohr, Salice et al. 2016). However, for obtaining more realism in respect to spatial-temporal dimensions and biodiversity, we require not only the identification of priority species, but also the spatial-temporal dimensions at which these species occur. Therefore, after identifying priority species, looking into the distribution patterns of these species can help to identify priority regions, that is, regions where these priority species are more abundant. These regions can assist in delivering realistic ranges of important landscape parameters (e.g. temperature, discharge, alkalinity) as input for environmental scenarios, enabling more realistic landscape level ERA (Franco, Price et al. 2016, Rico, Van den Brink et al. 2016). Additionally, these regions can become the focus of conservation and management efforts.

The two main objectives of the present study therefore are i) to construct models predicting the sensitivity of aquatic macroinvertebrates based on mode of action (MOA), traits and relatedness, and ii) to reveal spatial differences in the sensitivity of species composition assemblages by applying the developed models at the continental and national scale. The community composition of European freshwater ecoregions (ERs, based on Illies 1978) is used for the application of our models at the continental scale, while the reference database of the RIVPACS (River InVertebrate Prediction And Classification System) tool is used for river-type scale within the United Kingdom (Wright 1994). We conduct the first trait-based chemical sensitivity assessment of freshwater macroinvertebrate assemblages, extensively test the influence of spatial scale on sensitivity patterns, and provide key recommendations for its robust application in data-poor taxa.

3.2 Methods

The whole methodology of this study has been developed in R, a free software environment (R Core Team 2018). The R project, along with all scripts and data necessary to reproduce the models and figures performed in this study are available at Figshare (10.6084/m9.figshare.11294450) (Van den Berg 2019).

3.2.1 Modelling approach

We extracted toxicological data from Van den Berg et al. (Van den Berg, Baveco et al. 2019; original data from ECOTOX (USEPA, 2017)), which comprised Mode Specific Sensitivity (MSS) values for 36 and 32 macroinvertebrate genera towards baseline (narcosis) and AChE inhibiting toxicants respectively. Briefly, the MSS value represents the average relative

sensitivity of each species to a group of chemicals with the same MOA (original MOA classification from Barron, Lilavois et al. 2015), where an MSS value below zero indicates that the species is more sensitive than average, and an MSS value above zero indicates that the species is less sensitive than average. The MOAs narcosis and AChE inhibition were selected for this study, because they were the most data rich (Van den Berg, Baveco et al. 2019). Narcosis, also called baseline toxicity, is found toxic at similar internal concentration across all organisms (Wezel and Opperhuizen 1995, Escher and Hermens 2002) Therefore, differences in sensitivity for this MOA are expected to be mainly explained by traits related to toxicokinetics (i.e. uptake, biotransformation, and elimination). AChE inhibition is a more specific MOA, and therefore shows large differences in effect concentrations depending on taxonomic group (Van den Berg, Baveco et al. 2019). For this MOA we, therefore, expect a stronger phylogenetic signal. To justify a separate classification for the two MOAs, we made a correlation plot of the measured MSS values of species that were tested on both MOAs (Figure S3.7). The lack of a significant relationship between species sensitivity towards the two MOAs indicates that sensitivity towards these two MOAs is independent, and should therefore be evaluated separately.

The dataset from Van den Berg et al. (2019) also contained data on genus name, unique identifier (UID from the NCBI database, Benson, Karsch-Mizrachi et al. 2009, Sayers, Barrett et al. 2009), and traits (original data from Tachet, Richoux et al. 2000, Usseglio-Polatera, Bournaud et al. 2000). In this study, we added relatedness to this dataset by constructing a taxonomic tree, since detailed phylogenetic data was still largely unavailable or incoherent for most freshwater macroinvertebrates (we looked, for instance, in Genbank, Benson, Karsch-Mizrachi et al. 2009), and Guénard and Von der Ohe et al. (2014) have provided sufficient proof that taxonomic relatedness explains around the same amount of variation in species sensitivity as phylogenetic data when a wide taxonomic range is taken into consideration. This taxonomic tree is subsequently converted to Phylogenetic Eigenvector Maps (PEMs), from which species scores are extracted which subsequently serve as predictors of relatedness in model construction (Griffith and Peres-Neto 2006, Guénard, Legendre et al. 2013).

Constructing the taxonomic tree.

We constructed the taxonomic tree by extracting taxonomic data from the NCBI (National Centre for Biotechnology Information) database (Benson, Karsch-Mizrachi et al. 2009, Sayers, Barrett et al. 2009), followed by applying the *class2tree* function from the **taxize** package in R (version 0.9.3, Chamberlain and Szöcs 2013). Both the model species (for which we had sensitivity data available) and the target species (whose sensitivity we wanted to predict) were included in the tree. The simultaneous incorporation of both model and target species was necessary, because the PEM would change if the large number of target species would be added to the tree at a later point.

Phylogenetic eigenvector maps.

As descriptors of the taxonomic tree, phylogenetic eigenvectors were obtained from the PEM (see Guénard, Legendre et al. 2013 for details). PEMs work on a similar basis as principal component analysis (PCA; Legendre and Legendre 2012). Briefly, the eigenvectors of a PEM are obtained from a decomposition of the among-species covariance's and represent a set of candidate patterns of taxonomic variation of the response variables (i.e. the sensitivities to different chemicals). As is the case for a traditional PCA, this decomposition results in n – 1 eigenvectors (Legendre and Legendre 2012), where in our analysis n was the number of model species. The calculation of a PEM is obtained from both the structure of the taxonomic tree and from the dynamics of the (in our case) sensitivity evolution. The dynamics of the sensitivity evolution depends on the strength of a steepness parameter (parameter α; related to Pagels' parameter κ (Pagel 1999), where $\alpha = 1 - \kappa$). This parameter represents the relative evolution rate of the sensitivity to the MOA, takes values between 0 (natural evolution) and 1 (strong natural selection), and was in our study estimated from the known sensitivity of the model species. We constructed the PEMs with the **MPSEM** package (version 0.3-4, Guénard, Legendre et al. 2013, Guénard 2018).

Model construction.

For the narcosis dataset, two leverage points were discovered during the modelling process (Figure S3.1 and S3.2). Since we doubted the validity of these points (they were identical) and were unable to assess their validity (there was no data available on closely related species, and the reference was inaccessible), they were removed from the dataset, reducing the number of species for which toxicity data was available to 34. For the AChE inhibition dataset, only the 27 Arthropoda species present in the dataset were included in the analysis, because this MOA works in a more specific manner, making differences in MOA among different phyla more likely (Maltby, Blake et al. 2005). Eventually, 33 and 26 eigenvectors were included as taxonomic predictors for narcosis and AChE inhibition respectively (in the modelling process, taxonomic predictors were indicated with a 'V', see Figures S3.3 and S3.4 for examples of such predictors), and were added to the sensitivity and trait data. To reduce the number of predictors going into the final model building process (required due to memory limitations of the algorithm), an exhaustive search was performed using the *regsubsets* function from the **leaps** package (version 3.0, Lumley and Miller 2017). From this, traits or phylogenetic eigenvectors that were least frequently included in the best 1% of the models, ordered according to the Bayesian Information Criterion (BIC), were removed from the analysis. Next, an exhaustive regression was performed between the remaining predictors and the available MSS values, allowing a maximum of 4 predictors in the models. The best model was the model with the lowest AICc (Aikaike's Information Criterion with a correction for small sample size, Johnson and Omland 2004). The modelling exercise was repeated using only traits-, and a combination of traits- and taxonomic- predictors. We did not consider taxonomy-only models, because we

were primarily interested in obtaining more mechanistic understanding of the sensitivity process.

3.2.2 Predicting unknown taxa

The best model found for narcosis and the best model found for AChE inhibition were subsequently applied to the prediction of the sensitivity of species composition assemblages at two different spatial scales, continental and national. For the continental scale, the community composition of European freshwater ecoregions (ERs) was downloaded from https://www.freshwaterecology.info/ (Schmidt-Kloiber and Hering 2015). Although we realize that these data do not exactly resemble species assemblage data, it was the only dataset currently available at this spatial scale. For the national scale, the reference database of the RIVPACS tool was downloaded from the website of the Centre for Ecology and Hydrology (https://www.ceh.ac.uk/services/rivpacs-reference-database). The RIVPAC database was selected, because it is the only easily accessible database that provides detailed community level data at this spatial scale. The database contains macroinvertebrate assemblages at 685 reference sites, and was originally used to assess the ecological quality of UK rivers under the Water Framework Directive. To assess the ecological quality, the 685 sites have in an earlier study been grouped into 43 end groups based on biological and environmental variables (Davy-Bowker, Clarke et al. 2008). For descriptive summary purposes, these 43 end-groups were furthermore combined into 7 higher level super-groups (Davy-Bowker, Clarke et al. 2008, Table 3.1), such that these super-groups can be considered river-types at a relatively broad scale. In this study, we will use the super-groups to assess differences in species sensitivity on a river-type scale (**Table 3.1**).

The Tachet database was used as a source of traits data (Tachet, Richoux et al. 2000, Usseglio-Polatera, Bournaud et al. 2000). In order to make species-traits matching between the two community compositions (ERs and RIVPACS) and the Tachet database possible, the taxonomy of the three databases was aligned with the NCBI database using the **taxize** package (version 0.9.3, Chamberlain and Szöcs 2013). Species from the ER and RIVPACS communities could then be matched with traits from the Tachet database using the UIDs from the NCBI database. This matching was done at genus level. Since the traits in the Tachet database are coded using a fuzzy coding approach (describing a species by its affinity to several trait modalities, see Chevenet, Dolédec et al. 1994 for more information), a transformation was required before this data could be used. Continuous traits were transformed using a weighted averaging of the different trait modalities, whilst for factorial traits the modality for which the species had the highest affinity was selected (as in Van den Berg, Baveco et al. 2019).

At this point, taxonomic and trait data of all the target species (species for which we want to predict sensitivity) were complete, and PEM scores had to be added. To do this, the locations of the target species were extracted from the taxonomic tree, and subsequently transformed into PEM scores using the **MPSEM** package (version 0.3-4, Guénard, Legendre et al. 2013,

Guénard 2018). The PEM scores were then combined with the traits data, which allowed us to predict the sensitivity (MSS values) towards narcotic and AChE inhibiting chemicals using the two best models developed earlier.

The sensitivity of each ER or river type was determined by calculating the percentage of species with an MSS value below 0, comparable to (Hering, Schmidt-Kloiber et al. 2009). For RIVPACS, this was initially done both on abundance and presence-absence data, on the seasons spring, summer and autumn separately, and averaged over the three seasons. Eventually, we focused on presence-absence data averaged over the three seasons only, due to higher uncertainty (e.g. due to sampling error and seasonality) associated with the other data subsets. The results were projected on maps by colouring the ERs and river types according to the percentage of sensitivity species (MSS < 0) present. To construct the maps, we downloaded a map of the world from the Natural Earth website (https://www.naturalearthdata.com/downloads/10m-cultural-vectors/). The shape files for the ERs were obtained from the European Environment Agency (https://www.eea.europa.eu/data-and-maps/data/ecoregions-for-rivers-and-lakes), and their projection was transformed to match the projection of the world map using the *spTransform* function form the **sp** package (version 1.3-1, Pebesma and Bivand 2005). Coordinates of all the RIVPACS sites were available in the RIVPACS database.

3.2.3 Statistics

A Kruskal-Wallis Rank Sum Test was done to check if there were any statistically significant differences in sensitivity between ERs or RIVPAS groups. If this was true, multiple comparisons of all the groups were done with Kruskal Wallis using the *kruskal* function from the **agricolae** R package (version 1.2-8, Mendiburu 2017). Fisher's least significant difference criterion was used as a post-hoc test, and we used the Bonferroni correction as p-adjustment method.

Table 3.1. Division of the 685 reference sites into the 7 super-groups, along with a description of the dominant characteristics of the super-groups (taken from Davy-Bowker, Clarke et al. 2008).

RIVPACS super-group	N sites	Dominant characteristics
1	64	All in Scotland, mostly islands
2	148	Upland streams, mainly in Scotland and Northern England
3	169	Intermediate rivers, South-East Scotland, Wales, North and South-West England
4	48	Small steeper streams, within 13 km of source
5	115	Intermediate size lowland streams, including chalk, South-East England
6	84	Small lowland streams, including chalk, South-East England
7	57	Larger, lowland streams, South-East England, larger, finer sediments

3.3 Results

3.3.1 Sensitivity models

Incorporating taxonomic relatedness slightly improved the predictive capacity of models for invertebrate sensitivity towards narcotic and AChE inhibiting chemicals (higher adjusted R^2), compared to models without taxonomy (**Table 3.2**). Interestingly, the trait 'mode of respiration' was incorporated in the taxonomy & traits model of narcosis (Figure S3.3) and was also present in the traits-only model. For AChE inhibition, mode of respiration was included in the taxonomy & traits model (Figure S3.4), but not in the traits-only model. Considering the taxonomic predictors, V14, V2 and V4 were present in both the taxonomy-only and the taxonomy & traits model for narcosis. For AChE inhibition, the predictors V7 and V3 were present in both the taxonomy-only and the taxonomy & traits model.

Cross-validation of the model species resulted in the correct classification of 82% and 74% of the genera as sensitive or tolerant for respectively narcosis and AChE inhibiting chemicals (**Figure 3.1**). For narcosis, the Diptera *Paratanytarsus* and *Mochlonyx*, the Odonata *Ophiogompus*, the Ephemeroptera *Siphlonurus*, the Gastropoda *Aplexa*, and the Annelida *Chaetogaster* were misclassified (predicted on the wrong side of the zero line). For AChE inhibition, incorrect predictions were made in only two taxonomic groups, the Diptera *Glyptotendipes*, *Paratanytarsus*, *Tanytarsus*, and the Odonata *Anax*, *Crocothemis*, *Ophiogompus* and *Orthetrum*.

Table 3.2. Predictive models constructed for narcotic and AChE inhibiting chemicals, in- and excluding taxonomy. Taxonomic predictors are indicated with a V. See Figures S3.3 and S3.4 for a visualization of the predictors incorporated in the taxonomy & traits models.

MOA	Type of model	Model	Adj. R^2	p - value
Narcosis	Taxonomy & traits	MSS = -0.44 + 1.63 * V14 – 1.95 * V2 + 0.32 * respiration mode + 1.27 * V4	0.47	< 0.001
	Taxonomy-only	MSS = 0.16 + 1.66 * V4 + 1.64 * V14 + 1.16 * V5 – 1.14 * V2	0.42	< 0.001
	Traits-only	MSS = 0.04 – 0.25 * dispersal mode + 0.39 * respiration mode	0.20	0.011
AChE inhibition	Taxonomy & traits	MSS = 0.74 + 2.94 * V7 – 1.62 * V3 – 1.04 * V13 – 0.29 * respiration mode	0.62	< 0.001
	Taxonomy-only	MSS = 0.19 + 2.61 *V7 + 0.9 * V10 – 0.88 * V1 – 0.86 * V3	0.61	< 0.001
	Traits-only	MSS = 6.93 – 0.84 * life cycle duration – 1.13 * cycles per year– 0.17 * feeding mode – 0. 78 * temperature preferendum	0.4	0.004

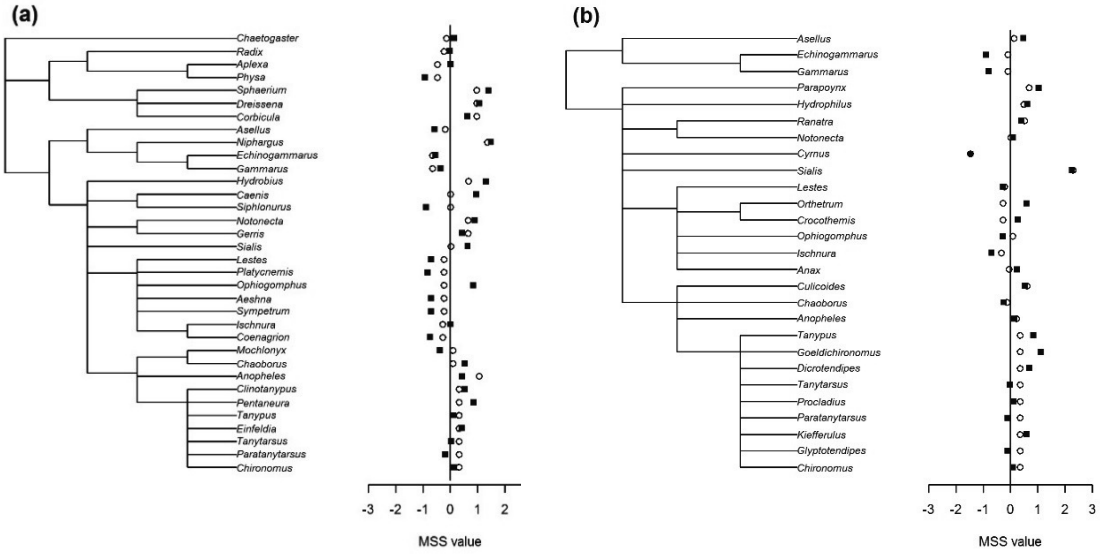


Figure 3.1. Observed MSS values (filled squares) and values predicted (unfilled circles) using traits and taxonomy according to the best models for (a) narcotic (b) and AChE inhibiting chemicals. Dendrograms show the taxonomic relationship between species according to class, family, order, and genus.

3.3.2 European freshwater ecoregions

Data availability.

For the ER communities, taxonomic data was available for 97% of the species, and covered four crustacean orders (Amphipoda, Anostraca, Decopoda, and Isopoda), and six insect orders (Coleoptera, Diptera, Ephemeroptera, Lepidoptera, Plecoptera and Trichoptera). Figure S3.5 shows the taxonomic composition of all ERs at the order level. For 19% of these species there was no or incomplete trait data available, leading to the exclusion of these species from our analysis. Of the remaining species, only around 5% had toxicity data available. We therefore had to predict the sensitivity of around 95% of the species for which no toxicity data was available using the taxonomy & traits models for narcosis and AChE inhibition.

Taxonomic pattern.

On the continental scale, 46 and 33% of the species were found sensitive (MSS < 0) towards narcotic and AChE inhibiting chemicals, respectively. For narcotic chemicals, 18 families contained only genera predicted as sensitive. Among these 18 families were all families belonging to the order of Isopoda (1 family), as well as a part of the Amphipoda (1 family), Plecoptera (6), and Trichoptera (10) families included in our study (Table S3.1). Five families contained both sensitive and tolerant genera. Four of these families belonged to the order of the Trichoptera, and one to the order of Lepidoptera. The remaining 25 families were predicted to only contain tolerant genera (MSS > 0), and included all of the families belonging to the order of Anostraca (1 family), Decapoda (5), Diptera (1), and Ephemeroptera (12), as well as the remaining Amphipoda (2 families), Plecoptera (1), and Trichoptera (3) families included in this study (Table S3.2).

For AChE inhibiting chemicals, there was little variation in sensitivity of the genera belonging to the same family, and the whole family was either predicted to contain only sensitive (MSS < 0) or only tolerant (MSS > 0) genera. All genera belonging to the order of the Trichoptera and all genera belonging to the family of the Gammaridae were predicted as sensitive (Table S3.3), while all other families included in this study were predicted to contain only tolerant genera (Table S3.4).

Geographical pattern.

For both MOAs, we noticed that the South of Europe (e.g. ER 1) has the highest proportion of sensitive species (MSS < 0), whilst Iceland (ER 19) is the ecoregion containing the lowest proportion of sensitive species (**Figure 3.2**). Central Europe (e.g. ER 14) contains the lowest percentages of sensitive species. ER 6 contains the largest percentage (57%) of species sensitive to narcotic chemicals, whilst ER 24 contains the largest percentage (45%) of species sensitive to AChE inhibiting chemicals.

When comparing the assigned sensitivity class of each ER for the two MOAs, we find that 8 of the 25 ERs were grouped into the same class for both MOAs (ER 1, 3, 5, 11, 18, 19, 21, 24, Figure S3.5). ER 2, 4, and 6 -10 were classified one or two classes lower for sensitivity towards AChE inhibiting chemicals compared to sensitivity towards narcotic chemicals, whilst the opposite was true for ER 12 -17, 20, 22, 23, and 25 (Figure S3.6).

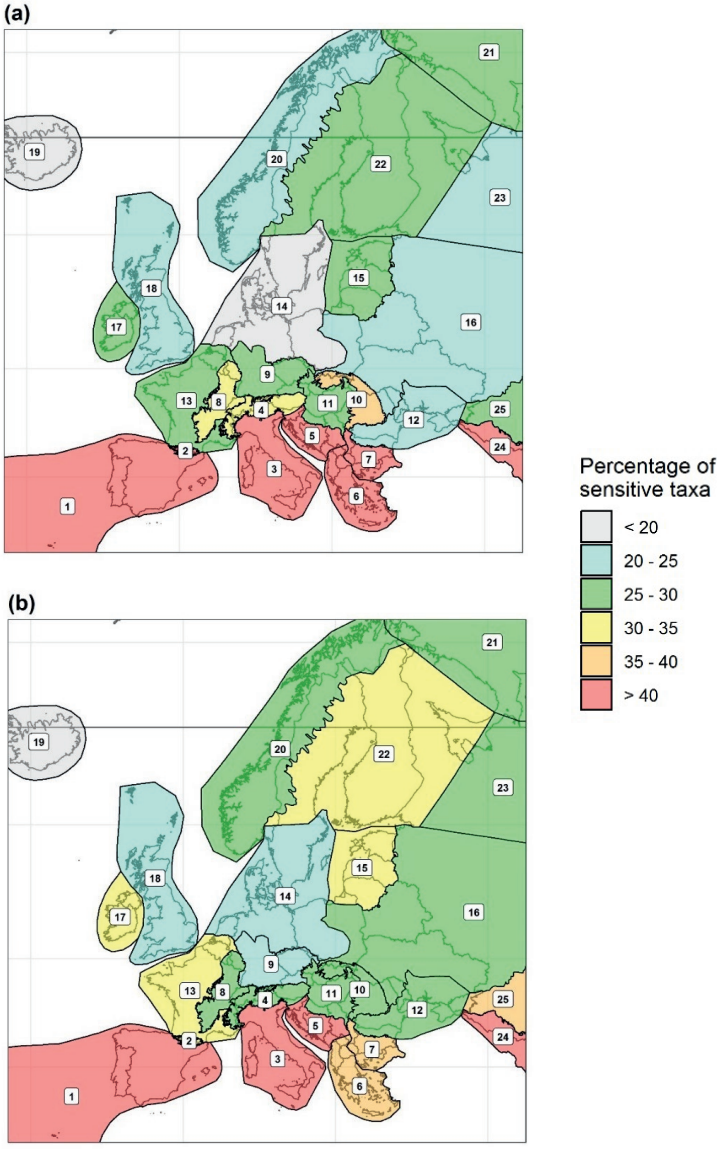


Figure 3.2. Percentage of sensitive taxa (MSS < 0) to narcotic (a) and AChE inhibiting (b) chemicals in European freshwater ecoregions. The numbers refer to the ecoregion number (ER 1 through ER 25).

3.3.3 RIVPACS river types

Data availability.

For the RIVPACS end-group communities, taxonomic data was available for 98% of the species. To ensure that model predictions did not trespass the taxonomic range on which the model was calibrated, any phylum that was not represented by one of the model species was removed from the analysis. Consequently, sensitivity towards narcotic chemicals was predicted for genera belonging to the phyla Annelida, Mollusca, and Arthropoda, whilst sensitivity towards AChE inhibiting chemicals was predicted only for Arthropoda. Coincidentally, in case of both datasets (Annelida, Mollusca, and Arthropoda, versus Arthropoda only), 34% of the species had no or incomplete traits data available, leading to the exclusions of these species from the analysis. Of the remaining species, less than 10% had toxicity data available. We therefore had to predict the sensitivity of 90% of the species for which no toxicity data was available using the taxonomy & traits models for narcosis and AChE inhibition.

Taxonomic pattern.

Within the UK, 38, and 25% of the species were found sensitive (MSS < 0) to narcotic and AChE inhibiting chemicals respectively. For narcotic chemicals, 37 families contained only genera predicted as sensitive, with an MSS value below zero. Among these 37 families were all families belonging to the order of Annelida (9 families), Isopoda (1), and Odonata (7), as well as a part of the Amphipoda (1), Plecoptera (6), Trichoptera (8), and Gastropoda (5) families included in our study (Table S3.5). Four families contained both sensitive and tolerant genera, all of them belonging to the order of Trichoptera. The 49 remaining families were predicted to only contain tolerant genera, with an MSS value above zero. Among them were all families belonging to the order of Arguloida (1 family), Coleoptera (7), Decapoda (1), Diptera (5), Ephemeroptera (9), Hemiptera (7), Lepidoptera (1), Megaloptera (1), Neuroptera (2), and Bivalvia (4), as well as the remaining Amphipoda (3), Plecoptera (1), Trichoptera (3), and Gastropoda (4) families (Table S3.6).

For AChE inhibiting chemicals, there was little variation in sensitivity of the genera belonging to the same family, and, as for the ER assemblages, the whole family was either predicted to only contain sensitive (MSS < 0) or tolerant (MSS > 0) genera. In total, 25 families contained genera that were all predicted as sensitive. This encompassed all families belonging to the order of Trichoptera (15 families), as well as a part of the Amphipoda (1), Diptera (2), Neuroptera (1), and Odonata (6) families (Table 3.8). The remaining 43 Arthropod families were predicted to only contain tolerant species, and included all Arguloida (1 family), Coleoptera (7), Decapoda (1), Ephemeroptera (9), Hemiptera (7), Isopoda (1), Lepidoptera (1), Megaloptera (1), and Plectopera (7), as well as the rest of the Amphipoda (3), Diptera (3), Neuroptera (1), and Odonata (1) families (Table S3.7).

Geographical pattern.

Considering the RIVPACS sites, geographical patterns show opposite results for the two MOAs (**Figure 3.3**). Regions containing more species sensitive towards narcotic chemicals were observed in the west and north of the UK, while regions containing more species sensitive towards AChE inhibiting chemicals were found in the south, south-west of the UK (**Figure 3.3**). RIVPACS sites located in small to intermediate lowland streams contained more sensitive species towards AChE inhibiting chemicals (super-groups 3, 4 and primarily 5, boxplots **Figure 3.3**), whilst for narcotic chemicals most sensitive species were found at sites located in upland rivers, mainly located in Scotland and Northern England (super-groups 1 and 2, boxplots **Figure 3.3**). For both MOAs, larger, lowland streams located in South-East England (super-group 7), contained the smallest percentage of sensitive species.

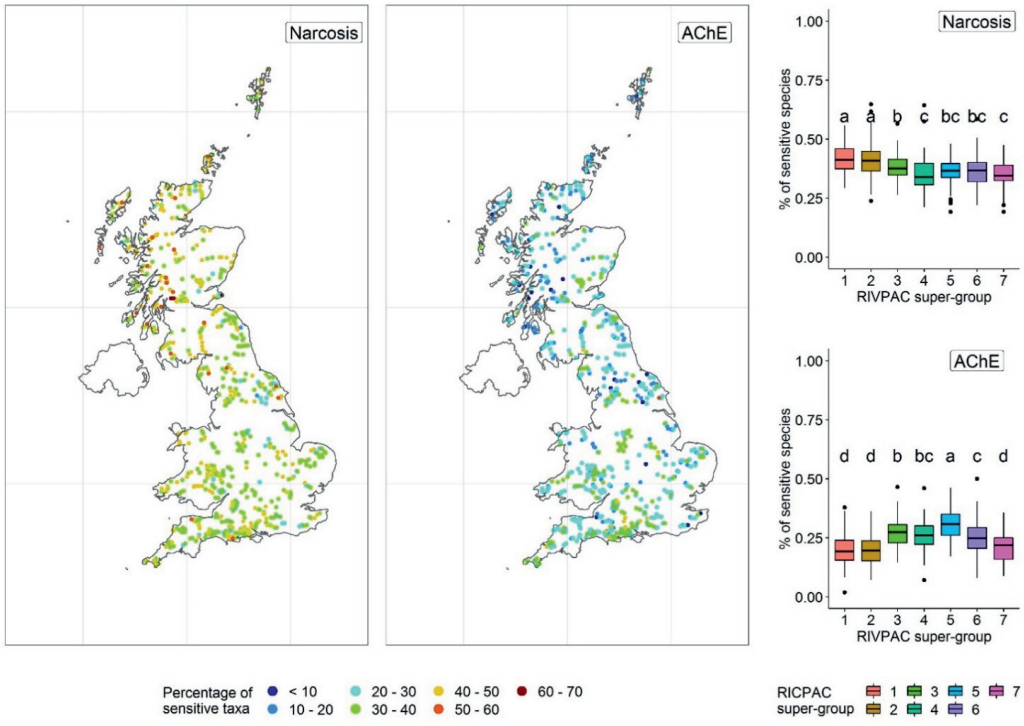


Figure 3.3. Map of the UK showing the percentage of sensitive taxa (MSS < 0) present at all RIVPACS sites, and boxplots of the percentage of sensitive species (MSS < 0) present in each RIVPACS super-group to narcotic and AChE inhibiting chemicals. Letters in boxplots indicate significant differences ($p < 0.05$).

3.4 Discussion

3.4.1 Traits and taxonomic predictor selection, and how this can be improved

For both MOAs, mode of respiration was selected as an important trait for explaining species sensitivity (**Table 3.2**). Several studies have investigated the relationship between respiration and AChE inhibiting chemicals before (Buchwalter, Jenkins et al. 2002, Rubach, Baird et al. 2010, Rubach, Baird et al. 2012, Rico and Van den Brink 2015, Van den Berg, Baveco et al. 2019), and have frequently found respiration important for determining species sensitivity, primarily due to an influence of respiration mode on uptake rates. The relationship between narcosis and respiration has been studied less, and there is to our knowledge only one study available that performed an analysis with narcotic chemicals (Van den Berg, Baveco et al. 2019). The result of that study closely aligns with ours, undoubtedly due to the large overlap in the data included in both studies.

We find that combining traits with taxonomic information results in models with increased predictive power, although only marginal (**Table 3.2**). Previous studies likewise emphasize the importance of complementing traits approaches with taxonomic approaches (Poff, Olden et al. 2006, Poteat, Jacobus et al. 2015, Pilière, Verberk et al. 2016). For example, Pilière and colleagues (2016) used boosted regression tree modelling to assess the environmental responses of single traits, orders and trait profile groups. They found that taxa belonging to the same trait profile group but to different orders showed different environmental responses. Similarly, they found that taxa belonging to the same order but to different trait profile groups showed different environmental responses (Pilière, Verberk et al. 2016). This indicates that unique information related to the evolutionary history was captured by the order of a taxon, whilst another part was captured by the trait set of a taxon. We find a similar result in our study, where the taxonomy-only model explaining sensitivity towards narcotic chemicals has an explanatory power of 0.42. This explanatory power increases to 0.46 when traits are included (**Table 3.2**). For AChE inhibition we see a similar result, although there the increase is only from 0.61 to 0.62 (**Table 3.2**). Although the increase of predictive power is only slight, the increase in mechanistic explanation is large, since the traits reveal mechanistic information regarding species sensitivity, and the taxonomic predictors point out taxa which show a different response to the chemical. The taxonomic predictors can thereby focus future research on finding the actual mechanisms that are different between these taxa. For this reason, both traits and taxonomy should be taken into consideration simultaneously for maximum benefit to risk assessment.

Although our models already show a good fit on the available data (**Table 3.2**), we anticipate that technological advances both in molecular and computational technologies will lead to an improvement of our models over time. Applying sophisticated molecular approaches can help with resolving the taxonomy of currently still problematic organism groups, for instance, by increasingly basing taxonomy on DNA markers, ideally replacing taxonomy completely by phylogenetics in due time (Hebert, Cywinska et al. 2003). Additionally, basing phylogenetic

trees on key target genes associated with Adverse Outcome Pathways (AOPs) might substantially improve phylogenetic predictive models for application in ecotoxicology (LaLone, Villeneuve et al. 2013). Furthermore, our models could improve with increased computing power. Due to memory limitations and the structure of currently existing model selection algorithms, we had to restrict the number of predictors going into the model selection process. However, since we maintain strict rules to avoid overfitting (e.g. the use of AICc as a model selection criterion and the use of a multivariate approach for the taxonomic predictors), it would be possible to add more predictors to the model without increasing the chance of overfitting.

3.4.2 Sensitivity patterns at European scale

At the continental scale, we predict that around half of the species are sensitive (MSS < 0) towards narcotic chemicals. This matches our expectations, since MSS is a relative value, and there is not any taxonomic group known that is particularly sensitive towards narcotic compounds (Escher and Hermens 2002). For AChE inhibiting chemicals we predict around one third of the arthropod species to be sensitive (MSS < 0). This is less than found in the sensitivity ranking of Rico and Van den Brink (Rico and Van den Brink 2015), where on average 70% of the Arthropoda were found sensitive towards AChE inhibiting chemicals (organophosphates and carbamates). However, this difference likely originates from the fact that Rico and Van den Brink (2015) also included non-arthropod species. Since MSS is a relative value, and arthropod species are the most sensitive group towards AChE inhibiting chemicals, including non-arthropod species will result in relatively more sensitive arthropod species.

Considering both MOAs, our predictions show that river basins in central Europe contain fewer sensitive species than those situated in the south (**Figure 3.2**). We reason that this results from, on the one hand, chemical exposure patterns before and during the period that Illies recorded the community composition of the ERs (Illies 1978), and on the other hand, from more ancient phylogeographical and ecological processes. Indeed, the pattern we find coincides with the emission pattern of multiple persistent organic contaminants commonly used in the 1960s, around the time when Illies was constructing his species database (Illies 1978). Chemicals like DDT (Dichloro-diphenyl-trichloroethane, Stemmler and Lammel 2009), lindane (Prevedouros, MacLeod et al. 2004), mercury (Pacyna, Pacyna et al. 2003), and PCDFs (polychlorinated dibenzofurans, Pacyna, Breivik et al. 2003) were more extensively used in central Europe, potentially reducing the occurrence of more sensitive species in those regions. However, we think that chemical exposure was not the main determinant for species composition, primarily because Moog and colleagues demonstrated that different ERs could always be differentiated from each other based on their community composition, even when heavily impacted by chemical stress (Moog, Schmidt-Kloiber et al. 2004). Therefore, we argue that the main cause for the geographical pattern we see lies in the phylogeography of Europe, in which extreme climatic events wipe out more sensitive species, and mountainous regions consecutively serve

as refugia and biodiversity hotspots (Rahbek, Borregaard et al. 2019, Rahbek, Borregaard et al. 2019). During the last ice age, glaciers covered the majority of northern Europe, forcing most species towards refugia present in southern Europe or to ice free parts of high mountain areas (e.g. Schmitt and Varga 2012). Indeed, there is a large overlap in biodiversity hotspots (Mittermeier, Myers et al. 1998, Médail and Quézel 1999, Rahbek, Borregaard et al. 2019) or so-called regions of large endemism (Deharveng, Dalens et al. 2000), with regions containing the highest percentage of sensitive species (**Figure 3.2**). Then after the last ice age, species recolonized northern Europe from these southern refugia, which is confirmed by the fact that almost all species occurring in northern European are also present in central and/or southern Europe (Hering, Schmidt-Kloiber et al. 2009). The relatively higher sensitivity of ER 22 and 15 (especially towards AChE inhibiting chemicals, **Figure 3.2**) can be explained due to migration of more sensitive species from Siberian refugia, e.g. located in the Ural mountains (Bernard, Heiser et al. 2011, Schmitt and Varga 2012).

3.4.3 Sensitivity patterns at UK scale

We see that certain biases in the underlying data are revealed in the sensitivity patterns we find for the UK. For instance, at a national scale, fewer species were considered sensitive compared to the continental scale, both towards narcotic and AChE inhibiting chemicals. We think this is caused by the interaction of two things. First, our models are biased in predicting entire families as sensitive or tolerant, in some cases resulting in entire phyla being predicted as sensitive or tolerant. Second, the RIVPACS communities are taxonomically uneven at genus level, the level we used to predict species sensitivity. Indeed, dipterans make up around 40% of all genera present which all are predicted to be tolerant towards the two MOAs. In this case, the taxonomic unevenness *at genus level* specifically, has a large influence on the percentage of species sensitive at the national scale. When we compare the ER and RIVPACS results at the family level, results between the two datasets are more consistent. For instance, for the ER dataset we predict that 33, 59, and 86% of respectively Amphipoda, Trichoptera, and Plecoptera families were sensitive towards narcotic compounds. This was 25, 53, and 86% of the families in the same orders in the RIVPACS dataset.

The geographical distribution of sensitive species throughout the United Kingdom is less pronounced than at a European level, although the opposing results of the RIPVAC super-groups towards the two MOAs studied is striking. This contradictory result corresponds with the study of Van den Berg et al. (2019), where an inclusive database approach reveals large differences in species sensitivity depending on MOA. Their study shows that AChE and narcosis are on opposing ends of a dendrogram clustered on a matrix of species sensitivity towards six diverse MOAs, indicating that AChE and narcosis show the largest differences in species sensitivity among all MOAs tested. Additionally, we found alternative explanations that could explain the contradicting geographical patterns we found for the two MOAs.

As an explanation for the geographical pattern for narcotic compounds, we find a large overlap between hotspots of sensitivity towards narcotic toxicants and conservation areas in the UK (e.g. with Special Areas of Conservation, Special Protection Areas, Sites of Special Scientific Interest, (Gaston, Charman et al. 2006)). It is known that protected areas serve as establishment centres, enabling the colonization of new regions by species that are shifting their geographical ranges (Thomas, Gillingham et al. 2012, Hiley, Bradbury et al. 2013). Although all RIVPACS sites are considered reference sites and have been selected because of low anthropogenic influence, our results show that whether these sites are included in or near to a conservational area leads to a higher support of sensitive species, likely due to an increased landscape and habitat heterogeneity.

As an explanation for the geographical pattern for AChE inhibiting compounds, the larger differences between the sensitivity of super-groups towards AChE inhibiting chemicals demonstrates that species sensitive towards AChE inhibition were more differentiated according to river type (i.e. the abiotic preferences of the species) than according to the availability of conservation areas. Additionally, the finding that the North to South pattern that we found at a European level was not noticeably present at the UK level is probably due to smaller differences in environmental factors (e.g. temperature, precipitation) when considering the UK only, compared to when the whole of Europe is considered.

3.4.4 Implications and outlook

Our analysis indicates that not only the taxonomic resolution of available trait databases is crucial, also the resolution of the (taxonomic) model is important. Additionally, we are confident that our models will improve in the near future, for instance by the replacement of the taxonomic tree with a phylogenetic tree based on validated biomarkers (for instance, as in Simões, Novais et al. 2019). In that case, the successful application of our suggested approach is mainly limited by access to raw biological data (e.g. species abundance), which is currently still problematic because governmental agencies provide ecological status information based on general indices rather than species counts. Providing access to raw data, along with clear metrics on the quality of that data, would foster our understanding of the links between anthropogenic stressors and populations or communities. Subsequently combining this effect data with chemical concentration data would be the next logical step, and would require chemical concentration data be made widely available by governmental agencies.

The current analysis provides an important new chapter in the development of environmental scenarios that can be used for the environmental risk assessment of chemicals at larger geographical scales (Franco, Price et al. 2016, Rico, Van den Brink et al. 2016). Our work is the first attempt to apply sensitivity models on community assemblage data previously grouped according to both biotic and abiotic parameters (e.g. invertebrate community composition, water depth, alkalinity and temperature, Davy-Bowker, Clarke et al. 2008). This combination of both biological and spatial data is required to successfully characterize exposure, effects and

recovery of aquatic non-target species under realistic worst-case conditions. Currently, mismatches exist between parameter values and spatial-temporal scales of ecological models used to predict potential effects of chemicals (Rico, Van den Brink et al. 2016). Our approach contributes to solving this mismatch by simultaneously incorporating biological and environmental factors.

In addition to this, the inclusion of traits in our models leads to an increased mechanistic understanding of cause-effect relationships, and allows for the application across wide biogeographical regions. This extrapolation enables, for instance, the comparison of ecological status across countries or regions that have so far remained unmonitored due to practical reasons (e.g. remote regions). Also, patterns across wide geographical scales can easily be compared with other studies to reveal regions where multiple stressors might be causing an effect simultaneously (e.g. Figure S3.6). Take, for instance, the potential impact of climate change on aquatic insects. Hering et al. (2009) show that southern European regions contain the highest fraction of species sensitive towards climate change. Since this largely overlaps with the regions we found to be most sensitive towards chemical stressors (**Figure 3.2**), there might be an increased overall effect on aquatic communities due to an unexpected interaction between climate change and chemical stress. In the north-east of Europe, a similar amplification effect may occur due to an overlap in regions with a relatively high chemical sensitivity (**Figure 3.2**), and predicted increased potential of harmful arthropod pest invasions (Bacon, Aebi et al. 2014).

Finally, our study demonstrates that sensitivity towards chemical stressors is spatially variable, and that although entire regions can be considered relatively tolerant, there might still be certain river reaches with a large percentage of sensitive species. Applied at relevant geographic scales, the methodology described in this study has demonstrated the potential to identify hotspots of sensitive species for given chemical classes. When applied to current risk assessment approaches, this will both increase the biological realism of assessments, and reduce the need for overly conservative assessment factors.

Acknowledgements

This work was supported by Unilever under the title ‘Ecological scenarios and models for use in risk assessment’ (ESMU). There are no conflicts of interest.

Supporting Information

Supporting Information to this article can be found on the following pages. The R project, along with all scripts and data necessary to reproduce the models and figures performed in this study are available at Figshare (10.6084/m9.figshare.11294450) (Van den Berg 2019).

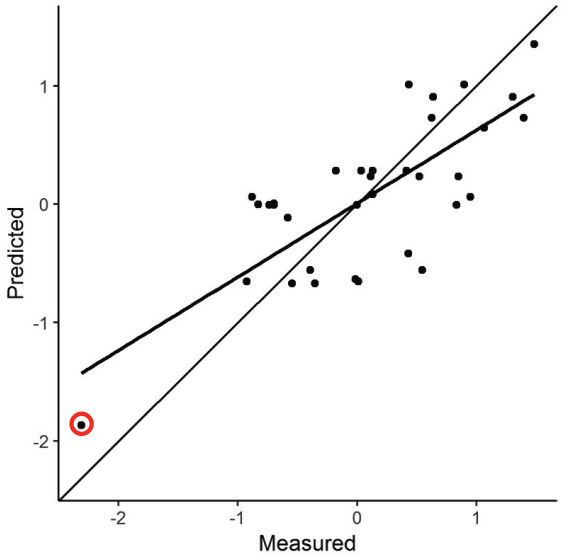


Figure S3.1. Measured MSS values plotted against predicted MSS values using the best model found for the narcosis dataset including the two leverage points. The red circle surrounds the two leverage points. The two leverage points lie exactly on the same spot.

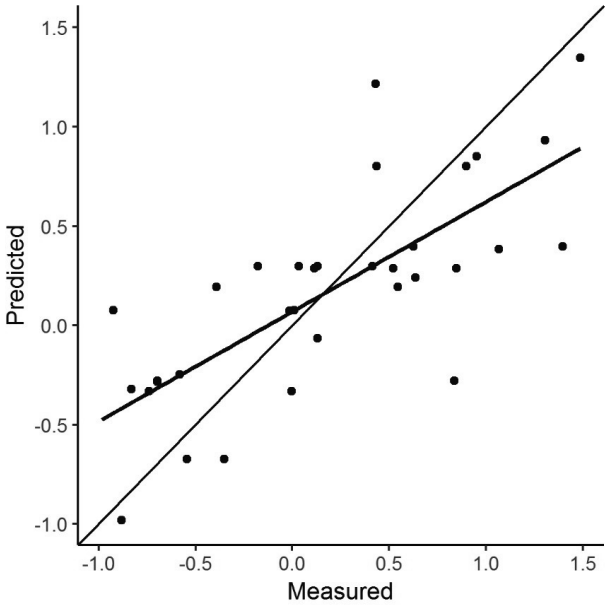


Figure S3.2. Measured MSS values plotted against predicted MSS values using the best model found for the narcosis dataset without the two leverage points.

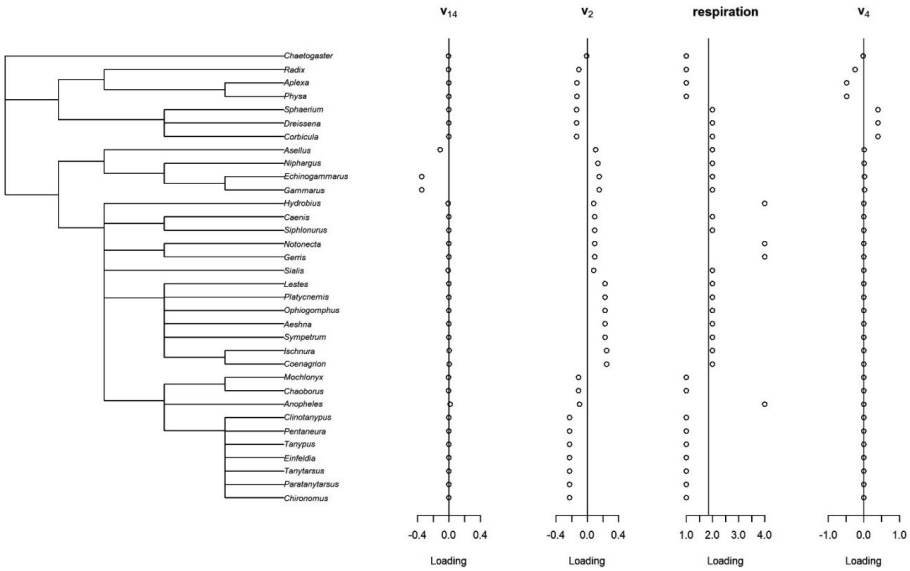


Figure S3.3. Visualization of the predictors incorporated in the best traits & taxonomy model for the narcosis dataset. The loadings of the taxonomic predictors and the trait values of the trait predictors are indicated for each genus on the taxonomic tree.

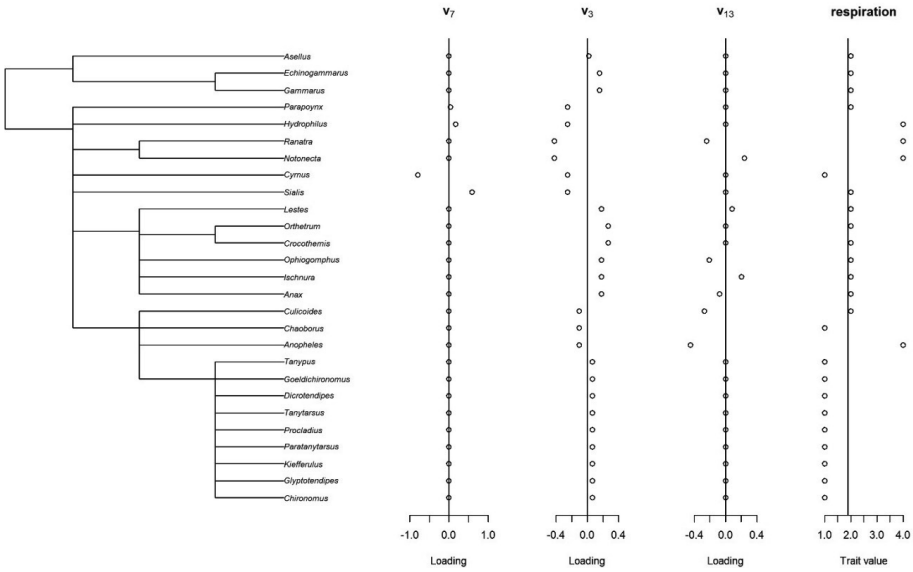


Figure S3.4. Visualization of the predictors incorporated in the best traits & taxonomy model for the AChE inhibtion dataset. The loadings of the taxonomic predictors and the trait values of the trait predictors are indicated for each genus on the taxonomic tree.

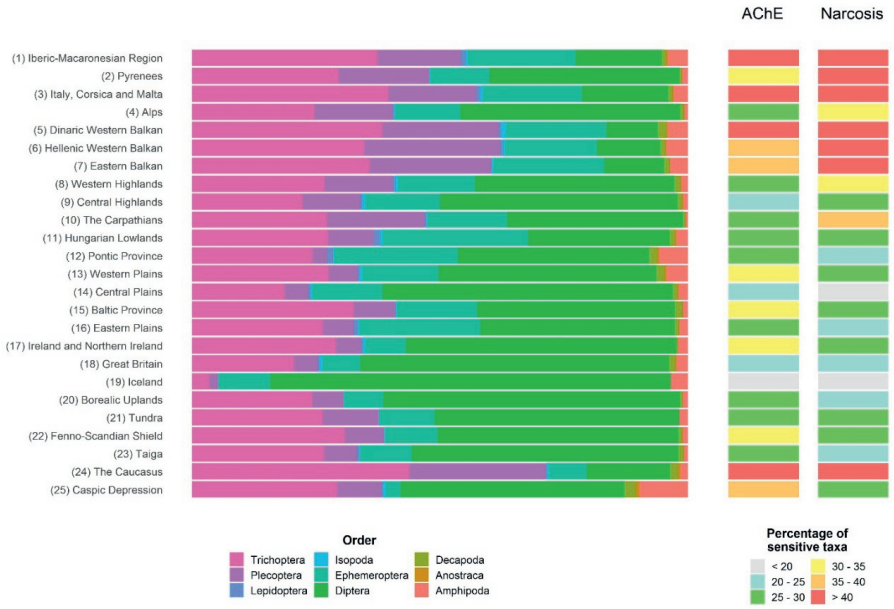


Figure S3.5. Taxonomic composition of ERs at order level.

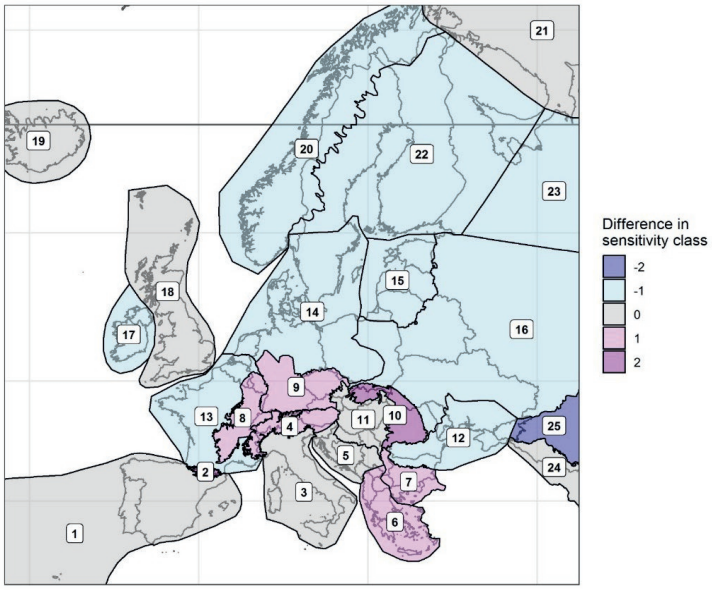


Figure S3.6. Relative sensitivity of the ERs towards the two MOAs tested.

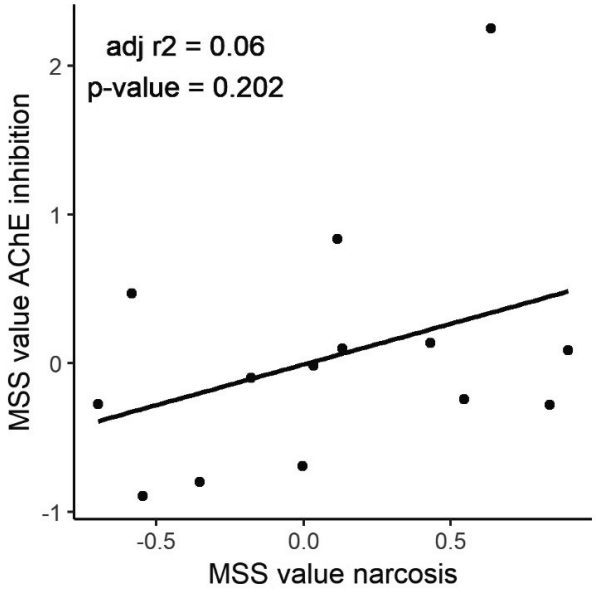


Figure S3.7. Correlation plot between the relative sensitivity (MSS) of species tested towards AChE inhibiting and narcotic chemicals. Each point indicates a different species.

Table S3.1. Number of genera from the European freshwater ecoregions predicted **sensitive** towards **narcotic** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. Shaded rows indicate that the family includes both sensitive and tolerant genera. The final row indicates the number of sensitive families per order.

	Amphipoda	Isopoda	Lepidoptera	Plecoptera	Trichoptera
Apataniidae	0	0	0	0	1
Asellidae	0	2	0	0	0
Beraeidae	0	0	0	0	2
Brachycentridae	0	0	0	0	2
Capniidae	0	0	0	2	0
Chloroperlidae	0	0	0	3	0
Crambidae	0	0	1	0	0
Ecnomidae	0	0	0	0	1
Gammaridae	2	0	0	0	0
Glossosomatidae	0	0	0	0	4
Helicopsychidae	0	0	0	0	1
Hydropsychidae	0	0	0	0	1
Hydroptilidae	0	0	0	0	8
Leptoceridae	0	0	0	0	4
Leuctridae	0	0	0	2	0
Limnephilidae	0	0	0	0	6
Nemouridae	0	0	0	4	0
Perlodidae	0	0	0	6	0
Philopotamidae	0	0	0	0	3
Polycentropodidae	0	0	0	0	6
Psychomyiidae	0	0	0	0	4
Taeniopterygidae	0	0	0	2	0
Uenoidae	0	0	0	0	1
Number of families	**1**	**1**	**1**	**6**	**14**

Table S3.2. Number of genera from the European freshwater ecoregions predicted **tolerant** towards **narcotic** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. Shaded rows indicate that the family includes both sensitive and tolerant genera. The final row indicates the number of sensitive families per order.

	Amphipoda	Anostraca	Decapoda	Diptera	Ephemeroptera	Lepidoptera	Plecoptera	Trichoptera
Ameletidae	0	0	0	0	1	0	0	0
Artemiidae	0	1	0	0	0	0	0	0
Astacidae	0	0	3	0	0	0	0	0
Atyidae	0	0	1	0	0	0	0	0
Baetidae	0	0	0	0	5	0	0	0
Beraeidae	0	0	0	0	0	0	0	1
Caenidae	0	0	0	0	2	0	0	0
Cambaridae	0	0	2	0	0	0	0	0
Chironomidae	0	0	0	122	0	0	0	0
Corophiidae	1	0	0	0	0	0	0	0
Crambidae	0	0	0	0	0	2	0	0
Crangonyctidae	1	0	0	0	0	0	0	0
Ephemerellidae	0	0	0	0	2	0	0	0
Ephemeridae	0	0	0	0	1	0	0	0
Goeridae	0	0	0	0	0	0	0	4
Heptageniidae	0	0	0	0	5	0	0	0
Hydropsychidae	0	0	0	0	0	0	0	2
Leptoceridae	0	0	0	0	0	0	0	4
Leptophlebiidae	0	0	0	0	6	0	0	0
Limnephilidae	0	0	0	0	0	0	0	2
Odontoceridae	0	0	0	0	0	0	0	1
Oligoneuriidae	0	0	0	0	1	0	0	0
Perlidae	0	0	0	0	0	0	2	0
Phryganeidae	0	0	0	0	0	0	0	2
Polymitarcyidae	0	0	0	0	1	0	0	0
Potamanthidae	0	0	0	0	1	0	0	0
Potamidae	0	0	1	0	0	0	0	0
Prosopistomatidae	0	0	0	0	1	0	0	0
Siphlonuridae	0	0	0	0	1	0	0	0
Varunidae	0	0	1	0	0	0	0	0
Number of families	**2**	**1**	**5**	**1**	**12**	**1**	**1**	**7**

Table S3.3. Number of genera from the European freshwater ecoregions predicted **sensitive** towards **AChE inhibiting** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. The final row indicates the number of sensitive families per order.

	Amphipoda	Trichoptera
Apataniidae	0	1
Beraeidae	0	3
Brachycentridae	0	2
Ecnomidae	0	1
Gammaridae	2	0
Glossosomatidae	0	4
Goeridae	0	4
Helicopsychidae	0	1
Hydropsychidae	0	3
Hydroptilidae	0	8
Leptoceridae	0	8
Limnephilidae	0	8
Odontoceridae	0	1
Philopotamidae	0	3
Phryganeidae	0	2
Polycentropodidae	0	6
Psychomyiidae	0	4
Uenoidae	0	1
Number of families	**1**	**17**

Table S3.4. Number of genera from the European freshwater ecoregions predicted **tolerant** towards **AChE inhibiting** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. The final row indicates the number of sensitive families per order.

	Amphipoda	Anostraca	Decapoda	Diptera	Ephemeroptera	Isopoda	Lepidoptera	Plecoptera
Ameletidae	0	0	0	0	1	0	0	0
Artemiidae	0	1	0	0	0	0	0	0
Asellidae	0	0	0	0	0	2	0	0
Astacidae	0	0	3	0	0	0	0	0
Atyidae	0	0	1	0	0	0	0	0
Baetidae	0	0	0	0	5	0	0	0
Caenidae	0	0	0	0	2	0	0	0
Cambaridae	0	0	2	0	0	0	0	0
Capniidae	0	0	0	0	0	0	0	2
Chironomidae	0	0	0	122	0	0	0	0
Chloroperlidae	0	0	0	0	0	0	0	3
Corophiidae	1	0	0	0	0	0	0	0
Crambidae	0	0	0	0	0	0	3	0
Crangonyctidae	1	0	0	0	0	0	0	0
Ephemerellidae	0	0	0	0	2	0	0	0
Ephemeridae	0	0	0	0	1	0	0	0
Heptageniidae	0	0	0	0	5	0	0	0
Leptophlebiidae	0	0	0	0	6	0	0	0
Leuctridae	0	0	0	0	0	0	0	2
Nemouridae	0	0	0	0	0	0	0	4
Oligoneuriidae	0	0	0	0	1	0	0	0
Perlidae	0	0	0	0	0	0	0	2
Perlodidae	0	0	0	0	0	0	0	6
Polymitarcyidae	0	0	0	0	1	0	0	0
Potamanthidae	0	0	0	0	1	0	0	0
Potamidae	0	0	1	0	0	0	0	0
Prosopistomatidae	0	0	0	0	1	0	0	0
Siphlonuridae	0	0	0	0	1	0	0	0
Taeniopterygidae	0	0	0	0	0	0	0	2
Varunidae	0	0	1	0	0	0	0	0
Number of families	**2**	**1**	**5**	**1**	**12**	**1**	**1**	**7**

Table S3.5. Number of genera from RIVPACS predicted **sensitive** towards **narcotic** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. Shaded rows indicate that the family includes both sensitive and tolerant genera. The final row indicates the number of sensitive families per order.

	Amphipoda	Gastropoda	Haplotaxida (Annelida)	Hirudinida (Annelida)	Isopoda	Lumbriculida (Annelida)	Odonata	Plecoptera	Trichoptera
Acroloxidae	0	1	0	0	0	0	0	0	0
Aeshnidae	0	0	0	0	0	0	2	0	0
Ancylidae	0	1	0	0	0	0	0	0	0
Apataniidae	0	0	0	0	0	0	0	0	1
Asellidae	0	0	0	0	1	0	0	0	0
Beraeidae	0	0	0	0	0	0	0	0	1
Brachycentridae	0	0	0	0	0	0	0	0	1
Calopterygidae	0	0	0	0	0	0	1	0	0
Capniidae	0	0	0	0	0	0	0	1	0
Chloroperlidae	0	0	0	0	0	0	0	1	0
Coenagrionidae	0	0	0	0	0	0	5	0	0
Cordulegastridae	0	0	0	0	0	0	1	0	0
Ecnomidae	0	0	0	0	0	0	0	0	1
Enchytraeidae	0	0	1	0	0	0	0	0	0
Erpobdellidae	0	0	0	3	0	0	0	0	0
Gammaridae	1	0	0	0	0	0	0	0	0
Glossiphoniidae	0	0	0	5	0	0	0	0	0
Glossosomatidae	0	0	0	0	0	0	0	0	2
Gomphidae	0	0	0	0	0	0	1	0	0
Haemopidae	0	0	0	1	0	0	0	0	0
Hydropsychidae	0	0	0	0	0	0	0	0	1
Hydroptilidae	0	0	0	0	0	0	0	0	4
Leptoceridae	0	0	0	0	0	0	0	0	2
Leuctridae	0	0	0	0	0	0	0	1	0
Libellulidae	0	0	0	0	0	0	2	0	0
Limnephilidae	0	0	0	0	0	0	0	0	6
Lumbricidae	0	0	1	0	0	0	0	0	0
Lumbriculidae	0	0	0	0	0	1	0	0	0
Lymnaeidae	0	4	0	0	0	0	0	0	0
Naididae	0	0	12	0	0	0	0	0	0
Nemouridae	0	0	0	0	0	0	0	4	0
Perlodidae	0	0	0	0	0	0	0	3	0
Philopotamidae	0	0	0	0	0	0	0	0	3
Physidae	0	3	0	0	0	0	0	0	0
Piscicolidae	0	0	0	1	0	0	0	0	0
Planorbidae	0	8	0	0	0	0	0	0	0

Platycnemididae	0	0	0	0	0	0	1	0	0
Polycentropodidae	0	0	0	0	0	0	0	0	5
Propappidae	0	0	1	0	0	0	0	0	0
Psychomyiidae	0	0	0	0	0	0	0	0	4
Taeniopterygidae	0	0	0	0	0	0	0	2	0
Number of families	1	5	4	4	1	1	7	6	12

Table S3.6. Number of genera from RIVPACS predicted **tolerant** towards **narcotic** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. Shaded rows indicate that the family includes both sensitive and tolerant genera. The final row indicates the number of sensitive families per order.

	Amphipoda	Arguloida	Coleoptera	Decapoda	Diptera	Ephemeroptera	Gastropoda	Hemiptera	Lepidoptera	Megaloptera	Neuroptera	Plecoptera	Trichoptera	Unionoida (Bivalvia)	Veneroida (Bivalvia)
Ameletidae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Argulidae	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Astacidae	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Athericidae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Baetidae	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
Beraeidae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Bithyniidae	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Caenidae	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
Chaoboridae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Chironomidae	0	0	0	0	87	0	0	0	0	0	0	0	0	0	0
Chrysomelidae	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Corixidae	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
Corophiidae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crambidae	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Crangonyctidae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Culicidae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Dixidae	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Dreissenidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Dytiscidae	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0
Ephemerellidae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Ephemeridae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Gerridae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Goeridae	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Helophoridae	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Heptageniidae	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
Hydraenidae	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Hydrobiidae	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Hydrochidae	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Hydrometridae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Hydrophilidae	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
Hydropsychidae	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Leptoceridae	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
Leptophlebiidae	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
Limnephilidae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Margaritiferidae	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Mesoveliidae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Nepidae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Niphargidae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Notonectidae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Odontoceridae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Osmylidae	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Perlidae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Phryganeidae	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Potamanthidae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Scirtidae	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Sialidae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Siphlonuridae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Sisyridae	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Sphaeriidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Unionidae	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
Valvatidae	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Veliidae	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Viviparidae	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Number of families	**3**	**1**	**7**	**1**	**5**	**9**	**4**	**7**	**1**	**1**	**2**	**1**	**7**	**2**	**2**

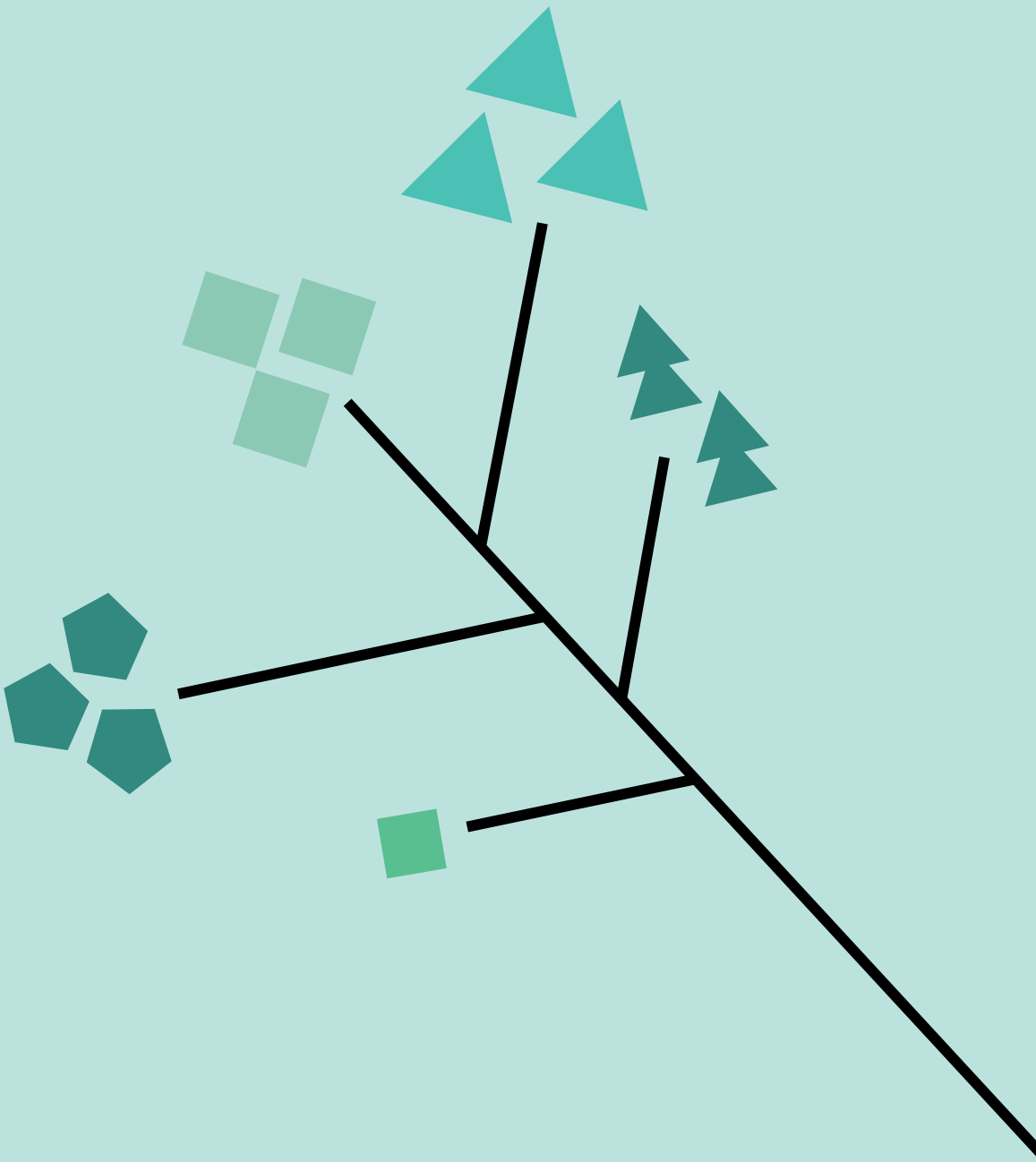
Table S3.7. Number of genera from RIVPACS predicted **tolerant** towards **AChE inhibiting** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. The final row indicates the number of sensitive families per order.

	Amphipoda	Arguloida	Coleoptera	Decapoda	Diptera	Ephemeroptera	Hemiptera	Isopoda	Lepidoptera	Megaloptera	Neuroptera	Odonata	Plecoptera
Ameletidae	0	0	0	0	0	1	0	0	0	0	0	0	0
Argulidae	0	1	0	0	0	0	0	0	0	0	0	0	0
Asellidae	0	0	0	0	0	0	0	1	0	0	0	0	0
Astacidae	0	0	0	1	0	0	0	0	0	0	0	0	0
Athericidae	0	0	0	0	1	0	0	0	0	0	0	0	0
Baetidae	0	0	0	0	0	4	0	0	0	0	0	0	0
Caenidae	0	0	0	0	0	2	0	0	0	0	0	0	0
Capniidae	0	0	0	0	0	0	0	0	0	0	0	0	1
Chironomidae	0	0	0	0	87	0	0	0	0	0	0	0	0
Chloroperlidae	0	0	0	0	0	0	0	0	0	0	0	0	1
Chrysomelidae	0	0	1	0	0	0	0	0	0	0	0	0	0
Corixidae	0	0	0	0	0	0	5	0	0	0	0	0	0
Corophiidae	1	0	0	0	0	0	0	0	0	0	0	0	0
Crambidae	0	0	0	0	0	0	0	0	1	0	0	0	0
Crangonyctidae	1	0	0	0	0	0	0	0	0	0	0	0	0
Culicidae	0	0	0	0	1	0	0	0	0	0	0	0	0
Dytiscidae	0	0	18	0	0	0	0	0	0	0	0	0	0
Ephemerellidae	0	0	0	0	0	1	0	0	0	0	0	0	0
Ephemeridae	0	0	0	0	0	1	0	0	0	0	0	0	0
Gerridae	0	0	0	0	0	0	1	0	0	0	0	0	0
Gomphidae	0	0	0	0	0	0	0	0	0	0	0	1	0
Helophoridae	0	0	1	0	0	0	0	0	0	0	0	0	0
Heptageniidae	0	0	0	0	0	3	0	0	0	0	0	0	0
Hydraenidae	0	0	3	0	0	0	0	0	0	0	0	0	0
Hydrochidae	0	0	1	0	0	0	0	0	0	0	0	0	0
Hydrometridae	0	0	0	0	0	0	1	0	0	0	0	0	0
Hydrophilidae	0	0	5	0	0	0	0	0	0	0	0	0	0
Leptophlebiidae	0	0	0	0	0	3	0	0	0	0	0	0	0
Leuctridae	0	0	0	0	0	0	0	0	0	0	0	0	1
Mesoveliidae	0	0	0	0	0	0	1	0	0	0	0	0	0
Nemouridae	0	0	0	0	0	0	0	0	0	0	0	0	4
Nepidae	0	0	0	0	0	0	1	0	0	0	0	0	0
Niphargidae	1	0	0	0	0	0	0	0	0	0	0	0	0
Notonectidae	0	0	0	0	0	0	1	0	0	0	0	0	0
Perlidae	0	0	0	0	0	0	0	0	0	0	0	0	2
Perlodidae	0	0	0	0	0	0	0	0	0	0	0	0	3

Potamanthidae	0	0	0	0	0	1	0	0	0	0	0	0	0
Scirtidae	0	0	3	0	0	0	0	0	0	0	0	0	0
Sialidae	0	0	0	0	0	0	0	0	0	1	0	0	0
Siphlonuridae	0	0	0	0	0	1	0	0	0	0	0	0	0
Sisyridae	0	0	0	0	0	0	0	0	0	0	1	0	0
Taeniopterygidae	0	0	0	0	0	0	0	0	0	0	0	0	2
Veliidae	0	0	0	0	0	0	1	0	0	0	0	0	0
Number of families	**3**	**1**	**7**	**1**	**3**	**9**	**7**	**1**	**1**	**1**	**1**	**1**	**7**

Table S3.8. Number of genera from RIVPACS predicted **sensitive** towards **AChE inhibiting** compounds. Each column indicates the order the genera belong to, whilst each row indicates the family. The final row indicates the number of sensitive families per order.

	Amphipoda	**Diptera**	**Neuroptera**	**Odonata**	**Trichoptera**
Aeshnidae	0	0	0	2	0
Apataniidae	0	0	0	0	1
Beraeidae	0	0	0	0	2
Brachycentridae	0	0	0	0	1
Calopterygidae	0	0	0	1	0
Chaoboridae	0	1	0	0	0
Coenagrionidae	0	0	0	5	0
Cordulegastridae	0	0	0	1	0
Dixidae	0	2	0	0	0
Ecnomidae	0	0	0	0	1
Gammaridae	1	0	0	0	0
Glossosomatidae	0	0	0	0	2
Goeridae	0	0	0	0	2
Hydropsychidae	0	0	0	0	3
Hydroptilidae	0	0	0	0	4
Leptoceridae	0	0	0	0	6
Libellulidae	0	0	0	2	0
Limnephilidae	0	0	0	0	7
Odontoceridae	0	0	0	0	1
Osmylidae	0	0	1	0	0
Philopotamidae	0	0	0	0	3
Phryganeidae	0	0	0	0	2
Platycnemididae	0	0	0	1	0
Polycentropodidae	0	0	0	0	5
Psychomyiidae	0	0	0	0	4
Number of families	**1**	**2**	**1**	**6**	**15**



CHAPTER 4

The combined potential of species-traits and mechanistic-effect models

Sanne J. P. Van den Berg, Paul J. Van den Brink, Andreas Focks

Abstract

Recently, mechanistic effect models have been suggested as an alternative to the statistical summaries (e.g. LC50s) that are currently being used to assess the potential risk of chemicals to the diversity of all living species. Examples of such quantitative mechanistic effect models are the toxicokinetic- toxicodynamic models of the General Unified Threshold models of Survival (GUTS) framework, which link external exposure and survival effects by describing the processes of uptake, biotransformation, elimination, damage and internal recovery. Parameterization of these models by means of traits enables prediction of standard sensitivity endpoints for a wide range of species and for multiple exposure patterns. In this study, the comparison of linear regressions between i) standard sensitivity endpoints (i.e. LC50 and EC50) and species traits, and between ii) GUTS parameters and species traits, shows that GUTS models parameterized on traits can approximate the sensitivity of freshwater arthropods towards the organophosphate chlorpyrifos, the pyrethroid lambda-cyhalothrin, and to a lesser extent, the neonicotinoid imidacloprid. We find that multiple quantitative links between traits and GUTS model parameters could be established, for instance, demonstrating that the uptake of chlorpyrifos is determined by lipid content and source of oxygen.

4.1 Introduction

A key challenge in environmental risk assessment (ERA) is to assess the potential risk of chemicals to the wide range of species present in the environment based on laboratory toxicity data derived from a limited number of species. Since it is impossible to test all possible species-chemical combinations in the lab, ERA must rely on modelling approaches to extrapolate known species sensitivities to realistic estimations of the effects of chemicals on the environment. One way to do this, is to make use of models which link species traits to a measure of sensitivity, so that in consequence the sensitivity of untested species can be predicted from the traits it possesses (e.g. Van den Berg, Baveco et al. 2019).

A decade ago, it has been proposed that the inclusion of species traits in ERA could provide a useful description of natural species assemblages and has the potential to replace classical taxonomic approaches (Baird, Rubach et al. 2008, Van den Brink, Alexander et al. 2011). Indeed, trait-based approaches have proven powerful in explaining differences in species occurrence patterns in aquatic ecosystems caused by natural variation, as well as by anthropogenic impacts (see Table 1 in Culp, Armanini et al. 2011, and the references therein). Trait-based approaches in ERA rely on the mechanistic relationships between biological and ecological characteristics and species sensitivity to different types of stressors, such as chemical stress, hydrological extremes, habitat deterioration, and nutrient excess or limitation. Regarding chemical stressors, previous studies have found that traits explained up to 77% of the difference in species sensitivity towards insecticides (Rubach, Baird et al. 2010, Rico and Van den Brink 2015). However, a more in-depth analysis of similar models constructed for a wider range of chemicals (e.g. baseline toxicants, neurotoxins) has proven that trait-based approaches are limited in their performance explaining standard sensitivity endpoints.(e.g. LC50, Van den Berg, Baveco et al. 2019) Whether this limit in model performance is due to the limitation of trait-based approaches in general, or due to the use of overall measures of sensitivity like EC50 to describe intrinsic species sensitivity instead of their underlying processes (Jager 2011), remains unclear.

In the past, quantitative mechanistic effect models have been suggested as an alternative for standard sensitivity endpoints like LC50 and EC50 (Jager, Heugens et al. 2006). Examples of such process-based models are the toxicokinetic (TK)- toxicodynamic (TD) models of the General Unified Threshold models of Survival (GUTS) framework, which describe the underlying processes of sensitivity. These models link external exposure and effects on survival by describing dynamically the process of TK (uptake, biotransformation, and elimination) and TD (damage/hazard, internal recovery and thresholds), and when fully parameterized on experimental data, are able to predict standard sensitivity endpoints (e.g. LC50) for any given time point, as well as for any given exposure pattern (EFSA Panel on Plant Protection Products and their Residues, Ockleford et al. 2018). There is a growing recognition that the use of such mechanistic approaches in ERA may improve predictive and extrapolative power of simple

statistical models which are currently applied for the analysis of effects in toxicity testing. Indeed, GUTS models have recently been endorsed by EFSA for their possible use in regulatory risk assessment (EFSA Panel on Plant Protection Products and their Residues, Ockleford et al. 2018). One of the remaining limiting factors for the more frequent use of GUTS modelling is that parameters still need to be calibrated on experimental data for any new species-compound combination. If it would be possible to parameterize GUTS models in an alternative way, i.e. without performing experiments but instead by predicting its parameters using traits-based models, this would enable prediction of standard sensitivity endpoints for a wide species range and for multiple exposure patterns.

Rubach and colleagues (2011) constructed a framework connecting traits with mechanistic effect models (Figure 2 in Rubach, Ashauer et al. 2011), and argued that trait-parameterized mechanistic effect models can be used to predict intrinsic sensitivity. In a follow-up study, they applied the developed framework to a simplified one-compartment first-order kinetic model, dynamically describing the uptake (k_{in}) and elimination (k_{out}) rates of chlorpyrifos for fifteen aquatic invertebrate species (Rubach, Baird et al. 2012). Linear regressions between the fitted TK parameters and the traits of the tested species resulted in relationships stronger than those found between traits and standard sensitivity endpoints (e.g. LC50). However, only the TK part of the GUTS model was parameterized on traits in this study, and it therefore remains unclear whether GUTS models parameterized on traits can provide accurate predictions of standard sensitivity endpoints.

In our study, we hypothesize that all parameters of the GUTS models (describing both TK and TD) can be parameterized on traits, and can subsequently be used to predict standard sensitivity endpoints. We test this using toxicity data collected for 15, 13, and 10 freshwater arthropods respectively exposed to the organophosphate chlorpyrifos (CPF, Rubach, Crum et al. 2011), the pyrethroid lambda-cyhalothrin (LCY, Schroer, Belgers et al. 2004), and the neonicotinoid imidacloprid (IMI, Roessink, Merga et al. 2013). The comparison of linear regressions between i) standard sensitivity endpoints (LC50 and EC50) and species traits, and between ii) GUTS parameters and species traits, shows whether GUTS models parameterized on traits are indeed able to approximate LC50 or EC50 models parameterized on traits.

4.2 Methods

4.2.1 Data

This study required four types of data: i) data on standard sensitivity endpoints (i.e. LC50 and EC50), ii) data on survival and immobilization over time, iii) data on the chemical (internal and/or external) concentration over time, and iv) traits data.

For CPF, LC50 and EC50 values as well as data on survival and immobilization over time (after 0, 24, 48, 72, and 96 h of exposure) were available for 15 freshwater arthropods, including 8 insects (*Anax imperator, Chaoborus obscuripes, Cloeon dipterum, Notonecta maculata,*

Parapoynx stratiotata, Plea minutissima, Ranatra linearis, Sialis lutaria), 3 decapods (*Neocardinia denticulata, Procambarus* sp. (separate data for adult and juveniles)), the isopod *Asellus aquaticus*, the cladoceran *Daphnia magna*, and the amphipod *Gammarus pulex* (separate data for adults and juveniles) (Rubach, Crum et al. 2011). Of these 15 species, two species were flagged as outliers (*S. lutaria* and *N. denticulata*), and were removed from the analysis (see Figure S4.1 in the Supporting Information (SI)). Additionally, since our objective is to reveal cross-species differences in sensitivity, we removed the least sensitive life stage if a species had data on multiple life stages available, which was the case for *Procambarus* sp. and *G. pulex*. This was done to avoid any bias in the traits data, since many traits were equal or highly correlated between life stages. For all of the remaining species, both internal and external CPF concentrations over time were obtained from (Rubach, Ashauer et al. 2010).

For LCY, LC50 and EC50 values as well as data on survival and immobilization over time (at least after 0, 48, and 96 h of exposure) were available for 13 freshwater arthropods, including 8 insects (*Chaoborus obscuripes, Cloeon dipterum, Caenis horaria, Erythromma viridulum, Notonecta glauca, Macropelopia* sp., *Sialis lutaria, Sigara striata*), two isopods (*Asellus aquaticus, Proasellus coxalis*), two cladocerans (*Daphnia galeata, Simocephalus vetulus*), and the amphipod *Gammarus pulex* (Schroer, Belgers et al. 2004). Regarding the chemical concentration over time, only external concentrations were available, and were obtained from (Schroer, Belgers et al. 2004).

For IMI, LC50 and EC50 values as well as survival and immobilization over time (after 0, 24, 48, 72, 96, of exposure, followed by observation of recovery in clean medium after 168, and 216 h) were available for 10 freshwater arthopods, including 8 insects (*Chaoborus obscuripes, Caenis horaria, Cloeon dipterum,* Limnephilidae, *Micronecta* spp., *Notonecta sp., Plea minutissima, Sialis lutaria*), the isopod *Asellus aquaticus*, and the amphipod *Gammarus pulex* (Roessink, Merga et al. 2013). Regarding the chemical concentration over time, only external concentrations were available, and were obtained from (Roessink, Merga et al. 2013).

Finally, traits data were taken from (Rubach, Baird et al. 2012). For the analysis twelve different traits were selected: size, water content, thickness of exoskeleton, lipid content, respiratory regulation, source of oxygen, mode of respiration, trophic relation, degree of sclerotization, Bauplan (shape of organism), life stage, and phylogeny (see **Table 4.1** for an overview of the traits, and their associated trait modalities). For CPF, trait data were available for all of the 11 species considered in this study. For LCY, the dataset was restricted to seven species, because trait data were unavailable for *C. horaria, E. viridulum, Macropelopia* sp., *S. striata, P. coxalis,* and *S. vetulus.* For IMI, the dataset was also restricted to seven species, because trait data were unavailable for *C. horaria,* Limnephilidae, and *Micronecta* spp..

Table 4.1. Description of traits included in the analysis, their modalities, abbreviations and type. Adjusted from (Rubach, Baird et al. 2012).

Abbreviation	Trait	Modality	Unit	Type of variable
Biovol	Size related	Biovolume	mm^3	Metric
SurfArea		Surface area (without gills)	mm^3	Metric
AVratio		Surface area/volume ratio	mm^{-1}	Metric
Length		Body length	mm	Metric
DryMass		Dry mass	mg/ind.	Metric
WatCont	Water content	Water content	%	Metric
ExoTh	Thickness of exoskeleton	Thickness of exoskeleton	mm	Metric
LipFW	Lipid content	% Lipid of wet weight	% wet weight	Metric
LipDW		% Lipid of dry weight	% dry weight	Metric
LipTot		Total lipid content	mg/ind.	Metric
ResConf	Respiratory regulation	Conformer	—	Binary
ResInt		Intermediate	—	Binary
ResReg		Regulator	—	Binary
SOatm	Source of oxygen	Atmospheric oxygen	—	Binary
SOdiss		Dissolved oxygen	—	Binary
ResMocut	Mode of respiration	Cutaneous	—	Binary
ResMosip		Siphon	—	Binary
ResMoCoG		Compressible gill	—	Binary
ResMoExG		External gills	—	Binary
ResMoInG		Internal gills	—	Binary
ResMoPig		Respiratory pigments	—	Binary
TroDetr	Trophic relation	Detritivore	—	Binary
TroHerb		Herbivore	—	Binary
TroCarn		Carnivore	—	Binary
TroOmni		Omnivore	—	Binary
SclPoor	Degree of sclerotization	Poor (<10%)	—	Binary
SclGood		Good (10-90%)	—	Binary
SclComp		Complete (>90%)	—	Binary
BauBox	Bauplan, shape of organism	Box shapes	—	Binary
BauCyl		Cylindroid	—	Binary
BauSphe		Spheres and ellipsoids	—	Binary
BauCone		Cones and half cones	—	Binary
Ladult	Life stage	Adult	—	Binary
Llarny		Larva/nymph	—	Binary
Ljuv		Juvenile	—	Binary
PhylRES	Phylogeny	Rank species (lowest rank = oldest)	—	Ordinal
PhylEQ		Rank taxon (lowest rank = oldest)	—	Ordinal

4.2.2 GUTS models

GUTS parameters resulted from fitting four versions of the GUTS model, the full (FULL) and the reduced (RED) model, combined with either the Stochastic Death (SD) or the Individual Tolerance (IT) assumption (RED-SD, RED-IT, FULL-SD, FULL-IT), to both survival (SUR) and immobility (IMMO) data. For a full description of the model we refer to (EFSA Panel on Plant Protection Products and their Residues, Ockleford et al. 2018) and see **Table 4.2** for an overview of the equations and parameters belonging to the four different GUTS models. SD models and IT models differ in the way they define the threshold of effects, and death above that threshold (**Table 4.2**). SD models have one value for the threshold of effects and an increased probability of death after exceeding it. IT models have a distribution of the threshold within the population, and death is instantaneous after exceeding the individual threshold. The FULL and RED version are distinguishable by the respective presence or absence of internal concentration data for model calibration (**Table 4.2**). Measurements of internal concentrations were only available for the experiments performed with CPF, and this is therefore the only chemical for which the full model could be fitted. We are aware that the assumptions and definitions of the equations and parameters associated with GUTS restrict its application to survival data (Jager, Albert et al. 2011). However, we apply the model to both survival and immobility data under the assumption that immobility is a pre-stage of mortality, assuming that recovery is impossible after four days of constant exposure (Ashauer, Agatz et al. 2011). Immobility data was corrected for recovery by fixing the number of immobile individuals from the timepoint where it first occurred to the rest of the experiment. All GUTS parameters were fitted using the openGUTS software version 1.0 (Jager 2019) implemented in Matlab (version R2018b). Since the openGUTS software only includes the RED model, the FULL model has been mimicked by transforming the measured external concentrations into simulated dynamic internal concentrations using the k_{in} and k_{out} parameters as obtained from (Rubach, Ashauer et al. 2010). These internal concentrations were then fed into the RED models in openGUTS, so that the other four parameters (k_R, b, m, and h_b for the SD model, and k_R, F_s, m and h_b for the IT model) could be optimized.

Table 4.2. Explanation of the equations, parameter symbols and their units used in this study.

Symbol	Explanation			Unit
Scaled damage	$\frac{dD_w(t)}{dt} = \boldsymbol{k_D}\left(C_w(t) - D_w(t)\right)$	Reduced	(1)	-
Internal concentration	$\frac{dC_i(t)}{dt} = \boldsymbol{k_{in}}C_w(t) - \boldsymbol{k_{out}}C_i(t)$	Full	(2)	-
Scaled internal damage	$\frac{dD_i(t)}{dt} = \boldsymbol{k_R}\left(C_i(t) - D_i(t)\right)$	Full	(3)	-
Hazard rate	$\frac{dH(t)}{dt} = \boldsymbol{b}\, max(0, D(t) - \boldsymbol{m})$	SD	(4)	-
Survival probability SD	$S_{SD}(t) = e^{-H(t) - \boldsymbol{h_b}t}$	SD	(5)	-
Survival probability IT	$S_{IT}(t) = \left(1 - F(t)\right)e^{-\boldsymbol{h_b}t}$, where $F(t) = \frac{1}{1+\left(\frac{\max_{0\le\tau\le t} D(\tau)}{\boldsymbol{m}}\right)^{-\boldsymbol{Fs}}}$	IT	(6) (7)	- -
$\boldsymbol{k_D}$	Dominant rate constant for the reduced model			day^{-1}
$\boldsymbol{k_{in}}$	Uptake rate constant for chemicals into the body			µg kg^{-1} day^{-1}
$\boldsymbol{k_{out}}$	Elimination rate constant for chemicals from the body			day^{-1}
$\boldsymbol{k_R}$	Damage repair rate constant			day^{-1}
$\boldsymbol{b_w}$	Killing rate constant, referenced to external concentration			L µg^{-1} day^{-1}
$\boldsymbol{b_i}$	Killing rate constant, referenced to internal concentration			kg µg^{-1} day^{-1}
$\boldsymbol{m_w}$	Median of the distribution of thresholds (IT), and threshold for effects (SD), referenced to external concentration			µg L^{-1}
$\boldsymbol{m_i}$	Median of the distribution of thresholds (IT), and threshold for effects (SD), referenced to internal concentration			µg kg^{-1}
Fs	Shape parameter for the distribution of thresholds			[-]
$\boldsymbol{h_b}$	Background hazard rate			day^{-1}

4.2.3 Trait-based models

The rest of the analysis was performed in the R environment (R Core Team 2020) and used some of the functions from the trait-based macroinvertebrate sensitivity modelling pipeline developed by Van den Berg et al. (2019). In short, after a pre-processing check for collinearity between traits, the remaining traits were further reduced by performing single linear regressions between all trait modalities and each of the endpoints separately. Subsequently, the trait modality which explained the largest variance in the respective endpoint went into an exhaustive multiple linear regression process. Eventually, the best model was selected from all constructed models based on the small sample unbiased Akaike's Information Criterion (AICc, Johnson and Omland 2004).

Linear regressions were performed between i) standard sensitivity endpoints (LC50 and EC50) and species traits, and between ii) GUTS parameters and species traits. Model performance was evaluated using model fit (R^2) and Leave-One-Out-Cross-Validation (LOOCV). For the latter, a mean squared prediction error (MSPE) and a prediction coefficient (P^2) were calculated as follows:

$$MPSE = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n} \quad (8)$$

$$P^2 = 1 - \frac{MSPE}{s_y^2} \quad (9)$$

where $\hat{y}_i$ is the estimated sensitivity or parameter value, y_i is the 'true' sensitivity or parameter value, and s_y^2 is the variance of the 'true' sensitivity or parameter values. When all predictions perfectly match observations, P^2 equals 1. Since the sizes of our datasets are small, further reducing this for the sake of cross-validation is arguable. When data is limited, a bad cross-validation result does not necessarily indicate an erroneous relationship, and literature might be available to provide support for the found relationship. However, good cross-validation results provide proof that the found relationship is consistent across species, and that the model is not performing well merely due to coincidence. Therefore, LOOCV results are used as a stringent measure to compare the mechanistic power of the two approaches tested in this study.

Finally, we used traits-predicted GUTS parameters as an input for the prediction of standard sensitivity endpoints, so that we can compare the predictions of LC50 and EC50 values using traits directly, or using traits for the parameterization of GUTS models. We compared the performance of the models by calculating the normalized root mean squared deviation (NRMSD) as measure for the distance of model prediction and observation.

4.3 Results and discussion

4.3.1 Relationship between traits and acute sensitivity

For CPF, the best models explained 88 and 50% of the total variation in EC50 and LC50 values respectively (**Figure 4.1**). The thickness of the exoskeleton (ExoTh) was found to have a positive relationship with both endpoints (Table S4.1), indicating that the thicker the exoskeleton of the organism, the less sensitive it is towards CPF. For EC50, the model additionally contained the traits life stage "juvenile" (Ljuv; negative (-) relationship) and being a respiratory regulator (ResReg; +) (Table S4.1). Although the negative sign of Ljuv indicates that species tested with juveniles were more sensitive than species tested with adults or larvae, we cannot conclusively say that juveniles are the most sensitive life stage, because our model did not include different life stages of the same species. That juveniles were found more sensitive than adults can, for instance, also result from a smaller body size, an effect found to explain sensitivity differences between adults and juveniles before (Gerritsen, van der Hoeven et al. 1998). The positive sign of respiratory regulation indicates that species which actively

regulate their respiration are less sensitive towards CPF. This relationship meets our expectations, and together with ExoTh indicates that uptake is an important process determining the variation in sensitivity towards CPF, a phenomenon well known from previous studies (Buchwalter, Jenkins et al. 2002, Rubach, Baird et al. 2012, Rotvit and Jacobsen 2013). Cross-validation results for both the LC50 and EC50 model were not good (P^2 value negative or close to zero, Table S4.1).

For LCY, the best models explained 75 and 41% of the total variation in EC50 and LC50 values respectively (**Figure 4.1**). The best models found for the two endpoints both contained only one trait (Table S4.1), ExoTh for the LC50 model, and phylogeny (PhylEQ) for the EC50 model. ExoTh indicates that uptake is an important process for LCY sensitivity, whilst PhylEQ indicates a phylogenetic signal in LCY sensitivity. From literature it is known that pyrethroids are rapidly taken up from the aqueous phase through the surface of the body due to the high octanol-water partitioning coefficient of these chemicals (it sorbs to lipids, Rasmussen, Wiberg-Larsen et al. 2013). Therefore, differences in the permeability of the organism caused by, for instance, differences in exoskeleton thickness, are likely to cause differences in species sensitivity. The importance of the dietary uptake of pyrethroids, however, is more unclear, since it is related to chemically and enzymatically mediated processes inside the digestive tract (Rasmussen, Wiberg-Larsen et al. 2013). Phylogeny is demonstrating to be a good metric for describing the differences in these enzymatically mediated sorption processes of the chemical. Besides a good model fit, the EC50 model for LCY also demonstrated good cross-validation results ($P^2 = 0.44$, Table S4.1), indicating that the phylogenetic signal of LCY sensitivity is well-conserved across species. Literature confirms that differences in arthropod sensitivity towards pyrethroids are dependent on mutations in sodium channels (Rinkevich, Du et al. 2013) or on alternative phylogenetic origins that are known to exist (Rinkevich, Su et al. 2012, Weston, Poynton et al. 2013).

For IMI, the best models explained 48 and 68% of the total variation in EC50 and LC50 values respectively (**Figure 4.1**), and like LCY, only one trait was included in the two models (Table S4.1). For LC50, this was the trait Length, confirming the well-known relationship between size and sensitivity (Poteat and Buchwalter 2014, Gergs, Kulkarni et al. 2015). For EC50 this was the trait PhylEQ, indicating a phylogenetic signal in IMI sensitivity. So far, the only empirical support of the presence of a phylogenetic signal in aquatic invertebrate sensitivity towards IMI, is that toxicity ranges in neonicotinoid sensitivity are larger among taxonomic groups than within taxonomic groups (Morrissey, Mineau et al. 2015). For bees, however, research explaining why they are so much more susceptible for IMI than any other insect is finding explanations in genomic pathways (e.g. in differences in the expression of cytochrome p450s, Manjon, Troczka et al. 2018, Beadle, Singh et al. 2019). Similar genomic pathways have been discovered in other terrestrial arthropods, like the hemipterans *Bemisia tabaci* (Karunker, Benting et al. 2008) and *Nilaparvata lugens* (Bass, Carvalho et al. 2011, Ding, Wen et al. 2013), and a phylogenetic signal in IMI sensitivity is therefore likely.

4.3.2 Relationship between traits and GUTS parameters

All GUTS model fits (available in S2) seemed reasonable, and the parameter values (available in S3) were therefore considered reliable for the construction of trait-based models.

For CPF, most of the models explaining GUTS parameters fitted on survival data had a similar (± 0.1) or better model fit (adjusted R^2) than the model explaining acute sensitivity (LC50, **Figure 4.1**). Only b_w in the reduced SD model for survival (SUR-RED-SD), and k_R and m_i in the SUR-FULL-IT model had a worse model fit than the LC50 model. For the GUTS parameters fitted on immobility data, all the RED-SD and FULL-SD parameters had a similar or lower model fit compared to the acute sensitivity model (EC50, **Figure 4.1**). Indeed, the EC50 model had a very good model fit, and was therefore difficult to surpass. However, for some parameters, such as k_{in} or m_i, fits of the SD parameters were almost reaching those of the EC50 model, and the F_s parameter of the IMMO-FULL-IT and the k_d parameter of the IMMO-RED-IT model still had a better fit than the EC50 model.

Regarding the traits that were included in the models for CPF, the parameters of the reduced models were most frequently explained by traits related to respiration (respiratory regulation (ResConf, ResReg), and mode of respiration (ResMoInG, ResMoPig)), but also the thickness of the exoskeleton (ExoTh) and the Bauplan (BauCone, BauCyl, BauBox) were regularly included in explaining some of the model parameters in the reduced models (Table S4.2). Regarding the parameters of the full model, k_{in} was explained by lipid content (LipTot) and source of oxygen (SOatm), whilst k_{out} was explained by phylogeny (PhylRES, Table S4.2). That k_{in} and k_{out} were consistently explained by the same traits across both the FULL-SD and FULL-IT model for both mortality and immobility data was expected, since the uptake (k_{in}) and elimination rate constants (k_{out}) were directly taken from (Rubach, Ashauer et al. 2010) and, therefore, their values did not change across the different full models. The identified explaining traits appear reasonable, since uptake is primarily determined by traits determining the uptake of the chemical, whereas elimination is determined by species-specific enzyme-based processes (see Kretschmann, Ashauer et al. 2011 for an example), for which phylogeny demonstrates to be a good proxy. However, an earlier study based on the same data (Rubach, Baird et al. 2012) identified additional and different traits (e.g. surface area) as explanatory variables for the uptake and elimination rate constants. These differences can be explained by two things. First, a different and more stringent algorithm was used in our study, which was primarily aiming to avoid overfitting of the models. Second, our models were constructed on the exact same data, but minus four data points: the adult life stage of *G. pulex* and *Procambarus* sp., and the leverage points of *S. lutaria*, *N. denticulata*.

The other parameters of the full models were most frequently explained by traits related to respiration (ResConf, ResInt) and shape of the organism (BauCone, BauCyl, BauSphe, Table S4.2), which is very similar to the traits used to explain the reduced model parameters. Indeed, when comparing the traits included in the full and the reduced model for CPF, we find that

similar traits, or opposite traits with an inverted relationship, were included in the parameter-model combinations that represent the same process in the full and in the reduced model. An opposite trait with the inverted relationship was, for instance, selected for the k_d parameter in the SUR-RED-SD model (BauCyl) compared to the k_R parameter in the SUR-FULL-SD model (-BauCone). The same is true for the m_w (RED-SD) - m_i (FULL-SD), the b_w (RED-SD) - b_i (FULL-SD), and the F_s (RED-IT) - F_s (FULL-IT) pairs, both for survival and immobility data. In general, an overlap in traits between the reduced and the full model confirms that those specific parts of the model explain the same or a related process in species sensitivity. The overlap in the traits composition of the k_R and k_d parameters provides additional information, namely that recovery appears to be the rate limiting step in CPF toxicity. If the dominant rate constant (k_d) would have been mainly determined by uptake or biotransformation processes, selected traits would have shown more overlap with traits showing correlations with k_{in} parameters, such as total lipid content (LipTot) and source of oxygen (Soatm) or k_{out}-related traits such as phylogeny (PhylRES). If, indeed, recovery is the rate limiting step, the fitting of two additional parameters in the full model (k_{in} and k_{out}) only assists in describing the bioaccumulation of CPF. This is confirmed by the finding that trophic relation (TroCarn) is used to describe k_R in the FULL-SD model, indicating that a difference in assimilation efficiency helps in explaining the differences in bioaccumulation. In general, carnivores have a different assimilation efficiency compared to other trophic levels (Hendriks, van der Linde et al. 2001), although studies focusing solely on invertebrates are lacking, making it difficult to hypothesize about the exact nature of this relationship.

For LCY, all models explaining GUTS parameters fitted on survival data (SUR-RED-SD, SUR-RED-IT) had a better or similar model fit than the model fitted on acute sensitivity data (LC50, **Figure 4.1**). For the GUTS parameters fitted on immobility data, the model fit was still good (lowest adj. R^2 of 0.43), but sometimes below the EC50 model (EC50, **Figure 4.1**). Regarding the traits that were included in the different models for LCY, both exoskeleton thickness (ExoTh) and shape of the organism (BauCyl) were frequently included in explaining different GUTS parameters (Table S4.3). ExoTh was twice found to explain the k_d parameter, both in the IMMO-RED-SD and the IMMO-RED-IT model, indicating that uptake and/or elimination was limited by the thickness of the exoskeleton. This matches its predictive power for explaining the LC50s, and the same reasoning applies.

For IMI, models explaining GUTS parameters fitted on immobility data performed better than the model fitted on acute sensitivity data (EC50) with the exception of one parameter (m_w, **Figure 4.1**). For the GUTS parameters fitted on survival data, two parameters had a better model fit (k_d and m_w in the SUR-RED-IT model), one parameter had a similar model fit (F_s in the SUR-RED-IT model), and three parameters had a lower model fit than the LC50 model (all parameters of the SUR-RED-SD model, **Figure 4.1**). The traits lipid content (LipDW), exoskeleton thickness (ExoTh), and shape of the organism (BauSphe) were most frequently

included in explaining the GUTS parameters for IMI (Table S4.4). This showed no overlap in the traits included in the LC50 and EC50 models.

Regarding the LOOCV results for the GUTS parameters, comparison of the prediction coefficients (P^2, Figure S4.6) and the model fits (adjusted R^2, **Figure 4.1**) shows that although the adjusted R^2 of a model is high, this does not guarantee good cross-validation results. Take, for instance, the parameters of the IMMO-RED-SD model for LCY. Although the adjusted R^2 values for k_d, b_w and m_w are relatively high (0.43, 0.96 and 0.72, **Figure 4.1**), the measured-predicted plots of the parameters look bad (regression line between measured-predicted data is far from passing through 0, and its slope is close to 0 or negative, Figure S4.4). These bad cross-validation results are reflected in negative P^2 values (Figure S4.6). Indeed, when P^2 values are positive, the measured-predicted plot looks good. Take, for example, the cross-validation of the parameters of the IMMO-RED-IT model for CPF (Figure S4.2). All the parameters have a positive P^2 value (Figure S4.6), and additionally show a regression line between the measured and predicted data that is relatively close to passing through zero (Figure S4.2). The same pattern can be found for the other chemicals (Figures S4.2 – S4.5).

Our analysis does not show any relationship between the type (metric, binary, ordinal) and number of traits included in the models for the GUTS parameters and the accuracy of the predictions. Models with good cross-validation results can range from being based on a single binary (CPF-SURV-FULL-SD: $m_i \sim$ ResConf) to a combination of 4 binary traits (CPF-SURV-RED-IT: Fs ~ BauBox+Ljuv+ResConf+SclGood). A comparison of the number of traits included in the model and cross-validation results (P^2 values) indicate no clear relationship (Figure S4.6). Another general observation is that when we compare the accuracy of the model predictions (P^2 values) of the three chemicals, we find that CPF had 4 parameters of the reduced models with P^2 values > 0.5, whereas IMI had three, and LCY had none. Whether this is related to the different physicochemical properties (e.g., hydrophobicity), to the different modes of action, or to the lower number of observations available for LCY and IMI, is not clear from this study.

Due to the nature of the traits data used in this analysis, we expected a higher mechanistic match between traits and TK parameters than between traits and TD parameters. However, this does not seem to be reflected in the model performance. Comparing the average model fits (R^2 values) of the TK parameters with the average model fits of the TD parameters, we see no difference for CPF and IMI, and we even find a higher average model fit for the TD parameters for LCY. Regarding the accuracy of the model predictions (P^2 values) of the TK and TD parameters, we also do not see that the models explaining TK parameters perform better than the models explaining TD parameters. Nevertheless, although it is apparently possible to find strong relationships between TD parameters and traits, it is more difficult to construct empirically testable hypotheses on the current traits models explaining TD parameters. Consider, for instance, that the F_s parameter of the CPF-SURV-RED-IT model is explained by

a combination of the shape (BauBox) and life stage (Ljuv) of the organism, the respiratory regulation (ResConf), and the degree of sclerotization (SclGood, Table S4.2). Therefore, the mechanistic understanding of TD parameters remains restricted due to a lack of more appropriate traits describing, for instance, the presence, absence, or distribution of target sites in the different organisms (Rubach, Ashauer et al. 2011).

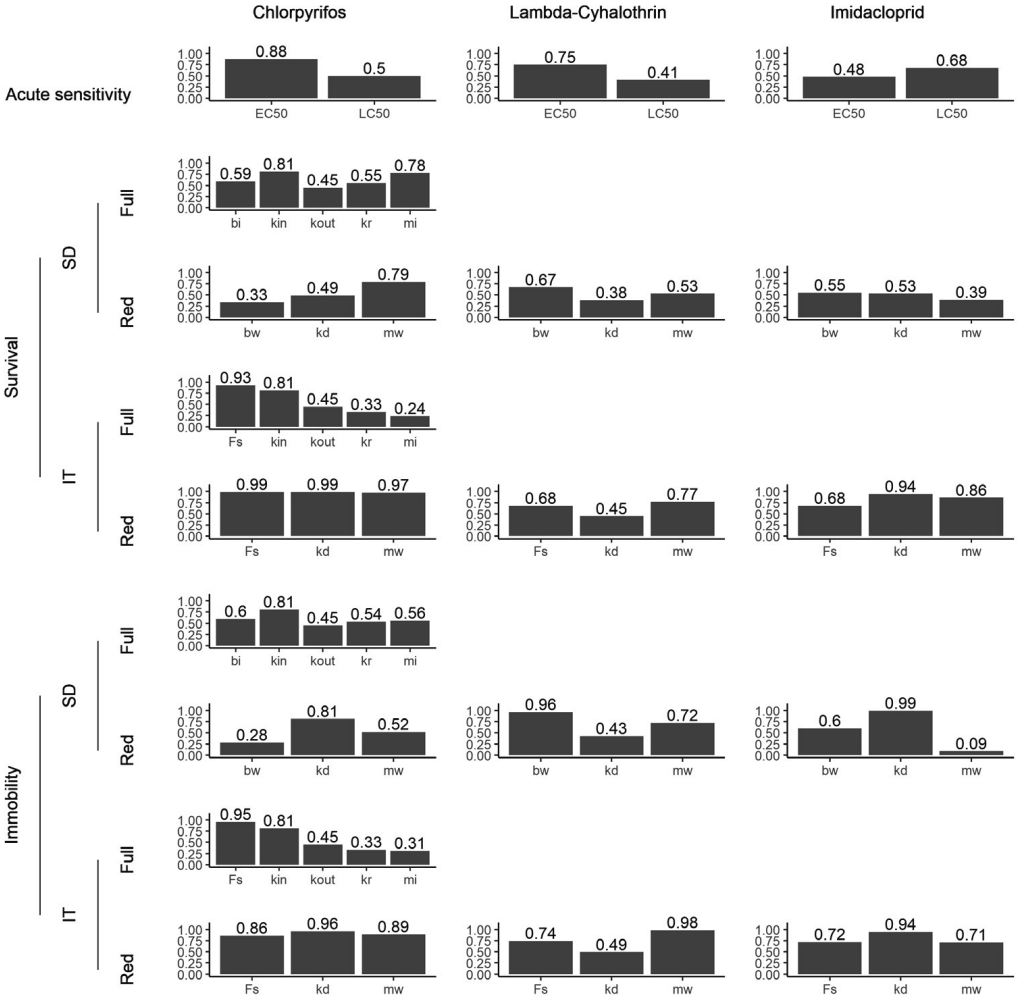


Figure 4.1. The model fits (adjusted R2) of the best models found with multiple linear regression for all endpoints under study.

4.3.3 Comparison of the direct and indirect prediction of standard sensitivity endpoints

The performance of the models directly predicting LC50 and EC50 values by means of traits turned out better than expected (**Figure 4.1 & 4.2**). With the exception of the EC50 model for CPF, all direct models only had a single trait incorporated in the best performing model, but still performed rather well (adjusted R^2 ranging between 0.41 and 0.75). Both the LC50 and EC50 models for CPF performed better than those of a similar analysis performed by Rubach and colleagues (Rubach, Baird et al. 2012). Our LC50 model also performed better than the single-traits LC50 model found for CPF by Ippolito et al. (2012). Possible reasons for the improved performance might include a better performance of the code or a superior data selection (e.g. removal of leverage points). For CPF and LCY, immobility data had a better model fit (smaller NRMSD) than mortality data, potentially indicating that sublethal effects are more closely related to sensitivity after short-term (4 days) exposure than lethal effects are.

Regarding the indirect prediction of standard sensitivity endpoints, where LC50 and EC50 values were estimated from predicted GUTS parameters, we find both promising and disappointing results when comparing them with the direct prediction (**Figure 4.2**). For CPF, the indirect prediction of LC50 values using the SUR-FULL-SD and SUR-RED-IT models performed better than the LC50 model, and the IMMO-RED-IT model performed almost as good as the EC50 model when the NRMSD is used as indicator for the accuracy of the model predictions.

When comparing the NRMSD for LCY, the IMMO-RED-SD model performs better than the EC50 model and the IMMO-RED-IT model performs relatively well. However, both the SUR-RED-IT and the SUR-RED-SD model showed much higher deviations in model predictions than the LC50 model. Obviously, the higher complexity of the GUTS modelling (more parameters) did not result in better predictions. It can be speculated, that short-term exposure experiments do probably not provide enough information for a compound such as LCY, which is known for its extreme hydrophobicity (Rasmussen, Wiberg-Larsen et al. 2013). Prolonged duration of such toxicity tests could probably provide more information.

For IMI, all GUTS models performed worse than the LC50 and EC50 models, although the SUR-RED-IT and the IMMO-RED-SD models come close to the respective performances of the LC50 and EC50 model. The disappointing results of the GUTS models for IMI hint at a failed inclusion of some of the processes determining species sensitivity. A possible explanation is that IMI requires bioactivation, that is, a chemical biotransformation to a more toxic or bio-accumulative metabolite. If this is indeed the case, the TK part of the model should be additionally calibrated on the biotransformation product, for instance as described for diazinon (Ashauer, Hintermeister et al. 2010, Kretschmann, Ashauer et al. 2011). An alternative to an experimental analysis of the internal concentration dynamics and biotransformation would be a prolongation of the toxicity testing. Indeed, the study from which we obtained the short-term toxicity data also contained results from long-term toxicity tests performed with the same

species (Roessink, Merga et al. 2013). There was a large difference in the sensitivity ranking of the species after 4 or 21 days of exposure. Firstly, all species demonstrated an increased sensitivity after the longer exposure duration. Secondly, the order of the species, when ranked from least to most sensitive, changed dramatically according to the exposure duration. Considering immobilization, for instance, *S. lutaria* and *C. obscuripes* demonstrated a higher relative sensitivity after 21 days of exposure, whilst *G. pulex*, who was found relatively sensitive after 4 days of exposure, was ranked least sensitive after 21 days. Both factors indicate that a short-term exposure to IMI is not representative enough to reveal the real sensitivity process.

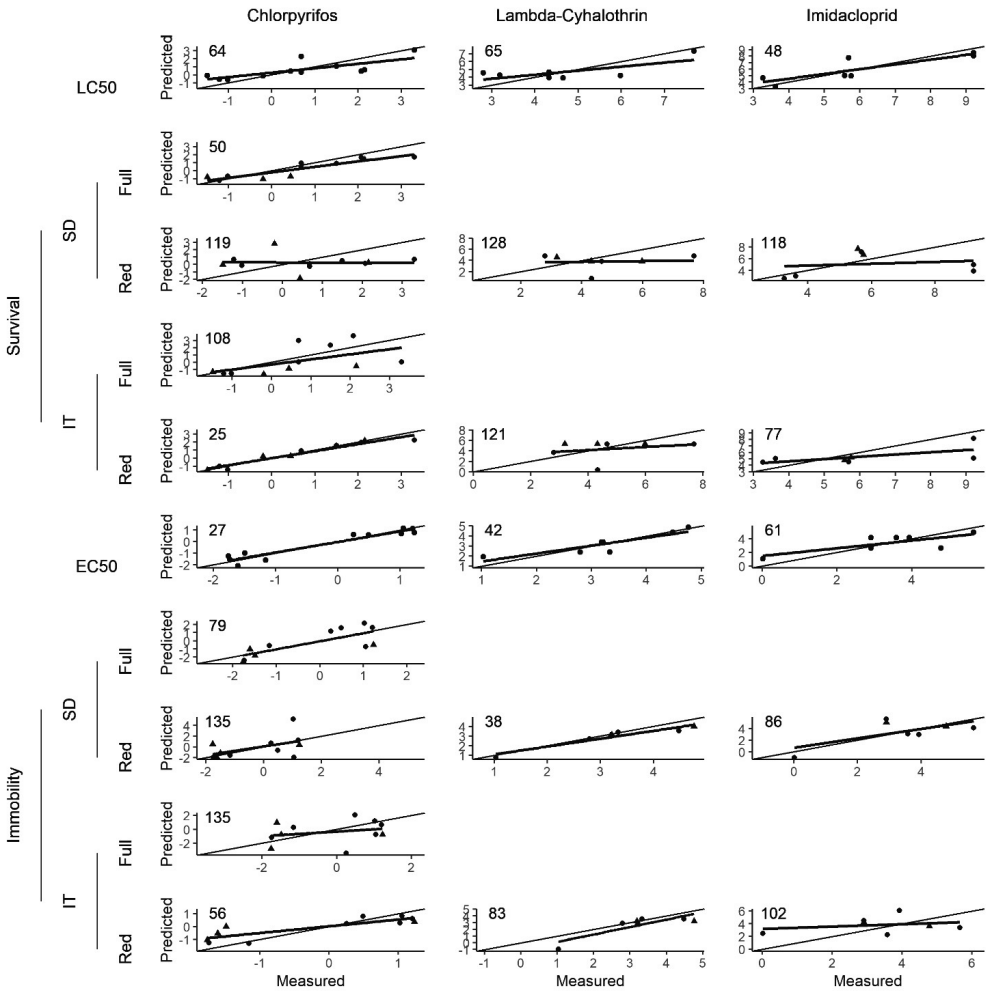


Figure 4.2. Measured versus predicted LC50 or EC50 values, after using trait-based prediction of standard sensitivity endpoints and GUTS parameters. The value inside each plot gives the NRMSD, indicating how accurate model predictions are.

In general, the SUR-RED-SD model never seems to give a good estimation of LC50 values, whilst the IMMO-RED-SD model works well for all three chemicals. To find a reason why the SUR-RED-SD model never seems to give a good estimation of LC50 values, we visually re-inspected all the GUTS model fits and the parameter space explored during the optimization process (S2). All model fits looked reasonable, and the frequency with which a parameter boundary was hit, or no clear optimal parameter value could be found, was similar for both the SD and the IT models. In consequence, immobility seems to be a better endpoint when aiming to create GUTS models based on species traits. This can be explained by the fact that immobilization is a more direct response to chemical stress than mortality, since organism death is the result of the interaction between multiple processes. In addition, immobility data have frequently been found as a more sensitive proxy for mortality (Maul, Brennan et al. 2008, Rubach, Crum et al. 2011, Raby, Nowierski et al. 2018). There are also critical discussions whether mortality is the best endpoint when the aim is to consider species sensitivity (Jager 2011, Jager 2012).

Considering the comparison of the full and the reduced models for CPF, we find that the full model performs better than the reduced model when the SD model is applied, whilst the opposite is true when the IT model is applied. This is true using both survival and immobility as endpoints. Internal concentrations were only available for CPF, so it is not possible to check this pattern for consistency across modes of action. In earlier GUTS validation studies (Nyman, Schirmer et al. 2012, Focks, Belgers et al. 2018), however, neither the RED-SD or the RED-IT models was superior to the other for the prediction of effects of untested time-variable exposure profiles, so that it appears that sometimes the RED-IT model and sometimes the RED-SD model is better in capturing the effect dynamics of a certain compound, which makes it less likely that the pattern that we find is of a general nature.

Considering the NRMSD values, our analysis does not show any relationship between the number of species included in the models and the accuracy of the predictions. Indeed, well-preforming models were found both for CPF, for sure one of the most data-rich examples of toxicity testing and GUTS modelling in literature, and LCY, a data-poor example that resulted in the second-best prediction of sensitivity out of all the models constructed (only the EC50 model for CPF had a smaller NRMSE). Therefore, we cannot claim that we simply need more species to improve these models. Of course, data on more species can reduce problems of collinearity between predictors, which can result in models with a better model performance, but it can also introduce new problems due to a broadening of the taxonomic scope, because the mode of action of a chemical might change depending on the taxonomic group (see Nendza and Muller 2000 for a more elaborate explanation).

So, if it is not necessarily more species that results in GUTS models with a better performance, then what is it? Considering the restrictedness of the data used in our analysis, our models have an impressive model performance, which can only be due to mechanistic relationships that we

have managed to exploit in our modelling pipeline. Unfortunately, our analysis is unable to determine the exact reason why some of the models worked better than others, but we hypothesize that the mechanistic overlap between species-traits and TKTD parameters has indeed been found in those models that performed well. For the models that did not outperform the standard sensitivity endpoints, the right traits were still lacking or removed during the collinearity check due to the small number of species for which data was available. Attempts to improve these models should focus on improving traits or alternative predictors that can mechanistically explain TD parameters.

4.3.4 Implications for future risk assessment

Our approach shows that although models with a good model fit can be found for standard sensitivity endpoints, models for GUTS parameters with an equal or even a slightly lower model fit are superior to standard sensitivity endpoints due to three reasons. First, the relationships between GUTS parameters and traits are easier to interpret mechanistically. This is especially true for TK parameters, although TD parameters still require the collection of alternative traits or other predictors that can mechanistically explain differences in TD parameters. Second, these mechanistic relationships do not only help us to understand the mechanism of sensitivity better, they also allow the formulation of new hypotheses regarding the mode of action of the chemical. Ashauer and colleagues demonstrate that TD parameters cluster together when chemicals belong to the same mode of action, and hypothesize how this, in combination with TK parameters extrapolated from other species using read-across, can enable the prediction of the toxicity of untested chemicals (Ashauer, O'Connor et al. 2015). Finally, the use of GUTS parameters allows the calculation of standard sensitivity endpoints using different exposure patterns. Previous work showed that fitted GUTS models can be used to predict the response of a number of species to a realistic exposure scenario (e.g. Focks, Belgers et al. 2018), which can assist in the more accurate development of SSDs based on representative species assemblages (e.g. Van den Brink, Buijert - de Gelder et al. 2019). Calibration of GUTS models without the necessity of performing experiments will open up new possibilities, even when the accuracy of trait-based GUTS modelling would be lower than that of experiment-based. In any case, without open access to the raw data underlying LC50 and EC50 values, this will be impossible.

The fitting of GUTS parameters requires more data than what is usually available in databases such as ECOTOX (USEPA 2017). Therefore, raw data from experiments, in case of GUTS that would be observations of survival and/or mobility over time need to be archived in its entirety, instead of summaries of data of inhomogeneous quality. Access to the full raw data would additionally enable transparency on experimental precision, another aspect remaining inadequately addressed by current approaches in ERA. Getting a firmer grip on uncertainty and variability associated with the execution of toxicity tests and species sensitivity can enlarge our capability to disentangle the two, and reveal the patterns in the noise.

Acknowledgements

This project has been funded by Unilever under the title 'Ecological Scenarios and Models for Use in chemical risk assessment' (ESMU).

Supporting information

(S1) PDF file: Supporting figures and tables. (S2) The GUTS model fits and explored parameter space. (S3) The R project, along with all scripts and data necessary to reproduce the models and figures performed in this study. S1 can be found on the following pages, whilst S2 and S3 are available at Figshare (10.6084/m9.figshare.11921025).

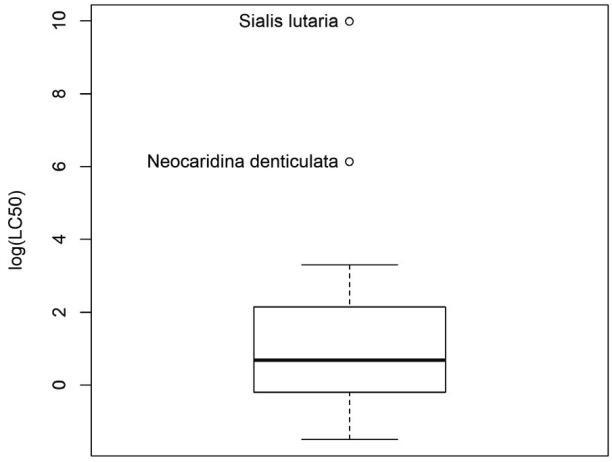


Figure S4.1. Outlier detection for the CPF dataset.

Table S4.1. Best trait-based models found for the classical sensitivity endpoints, LC50 and EC50, fitted on CPF, LCY and IMI data.

Chemical	Best model	Adjr. R^2	p value	LOOCV	P^2
CPF	LC50 ~ ExoTh	0.50	0.009	2.19	0.07
CPF	EC50 ~ ExoTh-Ljuv+ResReg	0.88	0.000	2.26	-0.32
LCY	LC50 ~ ExoTh	0.41	0.071	8.63	-2.09
LCY	EC50 ~ PhylEQ	0.75	0.008	0.84	0.44
IMI	LC50 ~ Length	0.68	0.014	5.57	0.02
IMI	EC50 ~ PhylEQ	0.48	0.052	6.26	-0.96

Table S4.2. Best trait-based models found for the TKTD parameters fitted on CPF data.

Endpoint	TK	TD	Best model	Adj. R^2	p value	LOOCV	P^2
Survival	Reduced	SD	kd ~ BauCyl	0.49	0.010	22.66	-0.59
Survival	Reduced	SD	mw ~ ExoTh - ResConf	0.79	0.001	10.62	0.51
Survival	Reduced	SD	bw ~ -ResConf	0.33	0.037	79.34	-14.89
Survival	Reduced	IT	kd ~ -BauCone-ExoTh+LipTot+Sodiss	0.99	0.000	2.16	-0.33
Survival	Reduced	IT	mw ~ ExoTh+LipTot+ResReg+WatCont	0.97	0.000	2.90	0.17
Survival	Reduced	IT	Fs ~ BauBox+Ljuv+ResConf+SclGood	0.99	0.000	0.06	0.93
Immobility	Reduced	SD	kd ~ -BauCone-Ladult+PhylRES	0.81	0.002	13.57	-1.17
Immobility	Reduced	SD	mw ~ PhylEQ-ResConf	0.52	0.022	148.69	-7.88
Immobility	Reduced	SD	bw ~ BauCone	0.28	0.053	7.51	-0.52
Immobility	Reduced	IT	kd ~ ExoTh+ResMoInG-ResReg	0.96	0.000	2.37	0.30
Immobility	Reduced	IT	mw ~ ExoTh+ResMoInG	0.89	0.000	2.01	0.62
Immobility	Reduced	IT	Fs ~ ResConf+ResMoPig-WatCont	0.86	0.001	0.20	0.58
Survival	Full	SD	kout ~ -PhylRES	0.45	0.014	0.78	-2.48
Survival	Full	SD	kr ~ -BauCone+TroCarn	0.55	0.017	8.60	-0.80
Survival	Full	SD	kin ~ -LipTot-Soatm	0.81	0.000	2.00	-0.40
Survival	Full	SD	mi ~ -ResConf	0.78	0.000	11.86	0.51
Survival	Full	SD	bi ~ ResInt+TroCarn	0.59	0.012	9.87	-1.77
Survival	Full	IT	kout ~ -PhylRES	0.45	0.014	0.78	-2.48
Survival	Full	IT	kr ~ BauCyl	0.33	0.037	20.48	-0.83
Survival	Full	IT	kin ~ -LipTot-Soatm	0.81	0.000	2.00	-0.40
Survival	Full	IT	mi ~ PhylRES	0.24	0.073	13.70	-1.30
Survival	Full	IT	Fs ~ -Llarny+ResConf+Sodiss-WatCont	0.93	0.000	2.38	-1.85
Immobility	Full	SD	kout ~ -PhylRES	0.45	0.014	0.78	-2.48
Immobility	Full	SD	kr ~ - BauCone-TroOmni	0.54	0.019	8.17	-1.93
Immobility	Full	SD	kin ~ -LipTot-Soatm	0.81	0.000	2.00	-0.40
Immobility	Full	SD	mi ~ PhylEQ-ResConf	0.56	0.016	124.58	-6.20
Immobility	Full	SD	bi ~ LipTot+TroCarn	0.60	0.010	5.96	-1.00
Immobility	Full	IT	kout ~ -PhylRES	0.45	0.014	0.78	-2.48
Immobility	Full	IT	kr ~ PhylEQ	0.33	0.039	46.36	-7.03
Immobility	Full	IT	kin ~ -LipTot-Soatm	0.81	0.000	2.00	-0.40
Immobility	Full	IT	mi ~ -BauSphe	0.31	0.045	26.84	-3.74
Immobility	Full	IT	Fs ~ - BauCyl+Ljuv+ResConf-Soatm	0.95	0.000	0.68	-0.08

Table S4.3. Best trait-based models found for the TKTD parameters fitted on LCY data.

Endpoint	TK	TD	Best model	Adj. R^2	p value	LOOCV	P^2
Survival	Reduced	SD	kd ~ -PhylEQ	0.38	0.082	48.15	-0.77
Survival	Reduced	SD	mw ~ -WatCont	0.53	0.039	13.81	0.03
Survival	Reduced	SD	bw ~ BauCyl	0.67	0.015	5.20	0.18
Survival	Reduced	IT	kd ~ Soatm	0.45	0.058	30.44	-1.10
Survival	Reduced	IT	mw ~ -BauCyl	0.77	0.006	3.55	0.18
Survival	Reduced	IT	Fs ~ WatCont	0.68	0.014	1.03	-1.26
Immobility	Reduced	SD	kd ~ -ExoTh	0.43	0.067	8.73	-0.67
Immobility	Reduced	SD	mw ~ -BauCyl	0.72	0.010	2.84	-0.19
Immobility	Reduced	SD	bw ~ -ExoTh+PhylEQ	0.96	0.001	2.05	-0.83
Immobility	Reduced	IT	kd ~ -ExoTh	0.49	0.047	29.02	-2.27
Immobility	Reduced	IT	mw ~ -BauCyl-ExoTh	0.98	0.000	8.88	-1.10
Immobility	Reduced	IT	Fs ~ SclPoor	0.74	0.008	0.19	0.36

Table S4.4. Best trait-based models found for the TKTD parameters fitted on IMI data.

Endpoint	TK	TD	lm	adjr2	p_value	LOOCV	P2
Survival	Reduced	SD	kd ~ ExoTh	0.53	0.04	58.45	-5.68
Survival	Reduced	SD	mw ~ BauSphe	0.39	0.08	33.19	-0.52
Survival	Reduced	SD	bw ~ LipFW	0.55	0.03	12.54	-0.41
Survival	Reduced	IT	kd ~ ExoTh	0.94	0.00	7.87	-4.01
Survival	Reduced	IT	mw ~ ExoTh	0.86	0.00	5.22	0.25
Survival	Reduced	IT	Fs ~ -BauSphe	0.68	0.01	1.24	-0.48
Immobility	Reduced	SD	kd ~ -Llarny-Soatm	0.99	0.00	5.74	0.68
Immobility	Reduced	SD	mw ~ WatCont	0.09	0.26	67.18	-2.02
Immobility	Reduced	SD	bw ~ LipFW	0.60	0.02	73.69	-5.52
Immobility	Reduced	IT	kd ~ -ResInt-Soatm	0.94	0.00	1.92	0.56
Immobility	Reduced	IT	mw ~ -LipFW	0.71	0.01	3.08	0.56
Immobility	Reduced	IT	Fs ~ -BauSphe	0.72	0.01	0.88	-0.11

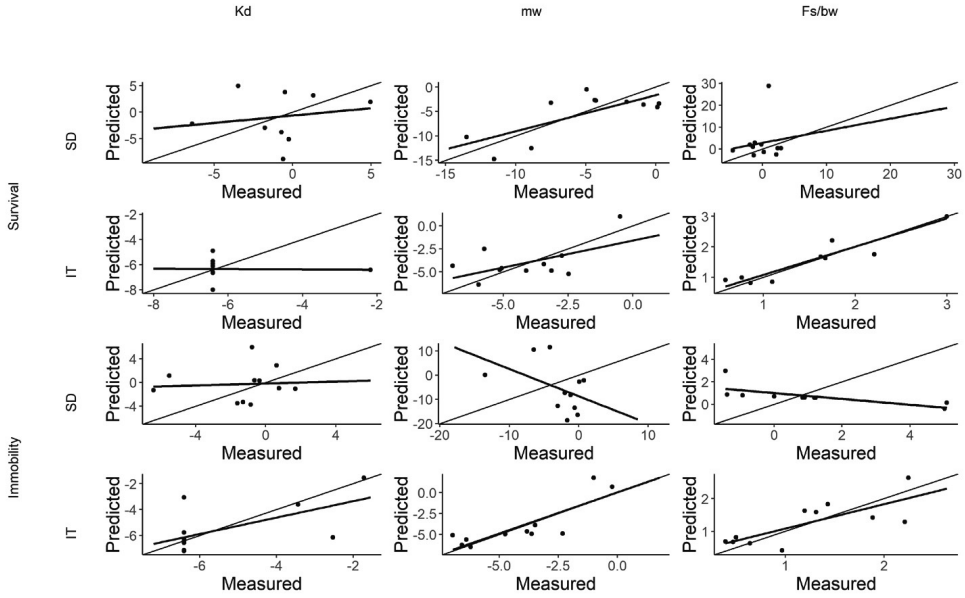


Figure S4.2. Measured-predicted plots of the TKTD parameters resulting from Leave-One-Out-Cross-Validation (LOOCV) performed for both survival and immobility and the RED-SD and RED-IT models for the chemical CPF.

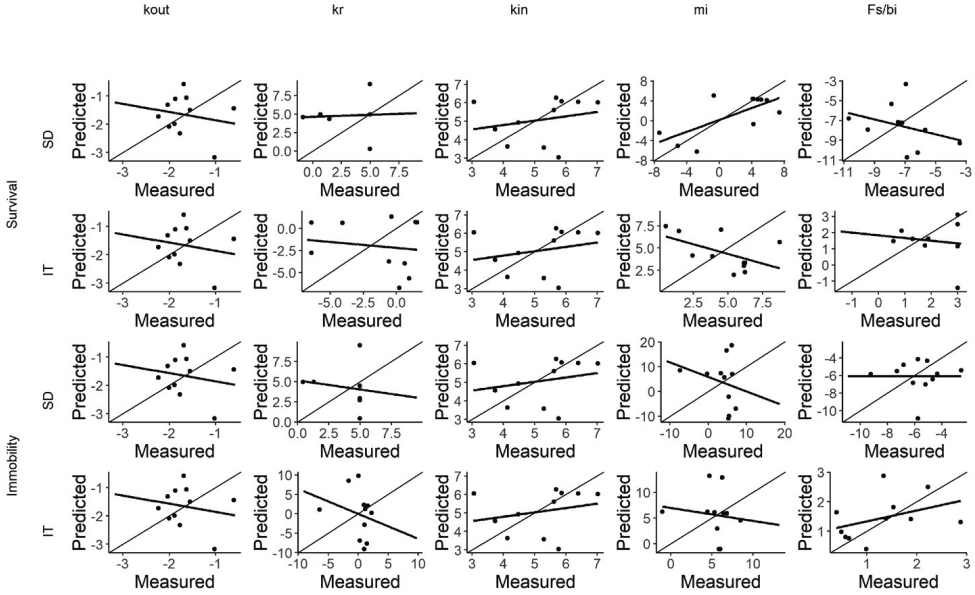


Figure S4.3. Measured-predicted plots of the TKTD parameters resulting from LOOCV performed for both survival and immobility and the FULL-SD and FULL-IT models for the chemical CPF.

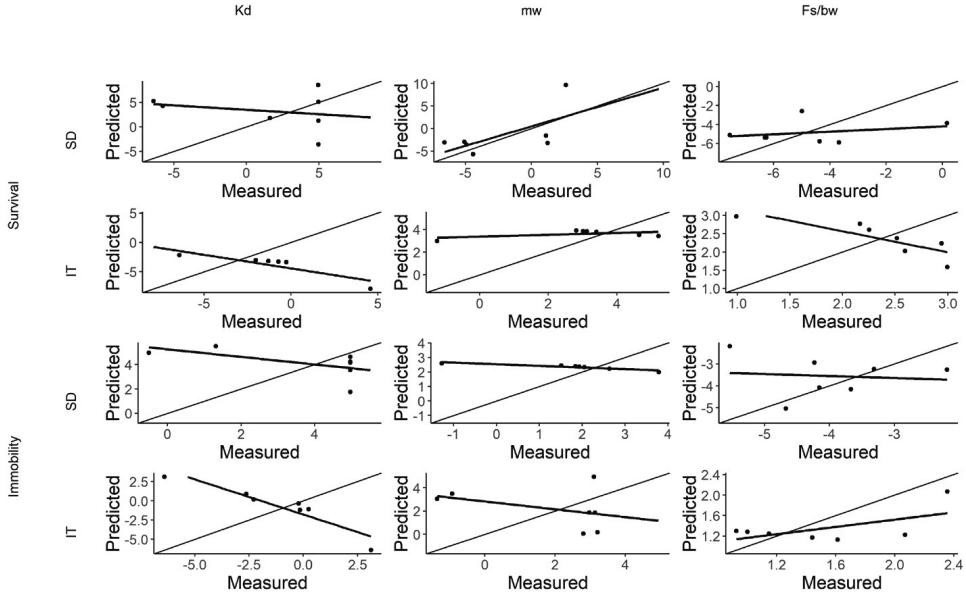


Figure S4.4. Measured-predicted plots of the TKTD parameters resulting from LOOCV performed for both survival and immobility and the RED-SD and RED-IT models for the chemical LCY.

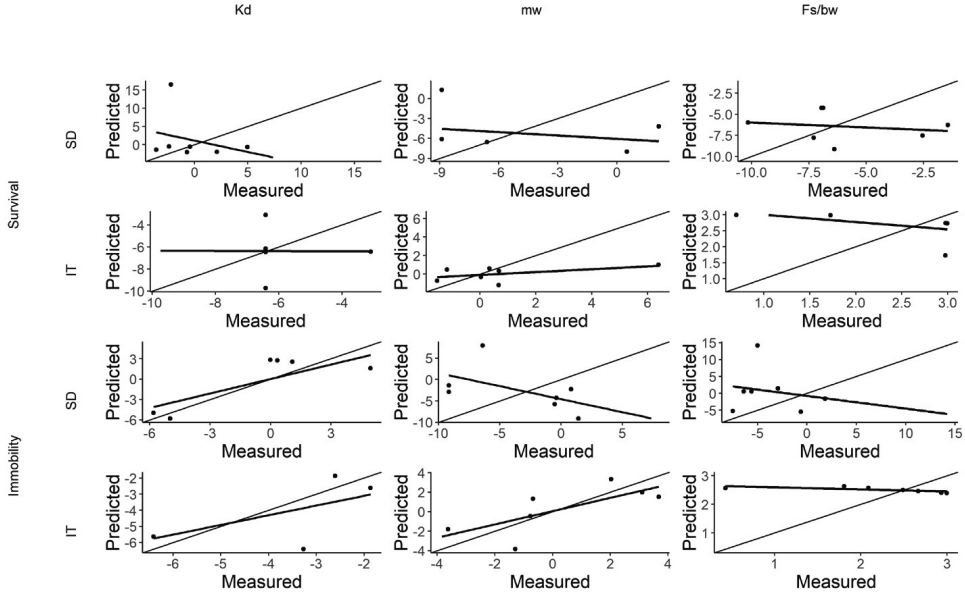


Figure S4.5. Measured-predicted plots of the TKTD parameters resulting from LOOCV performed for both survival and immobility and the RED-SD and RED-IT models for the chemical IMI.

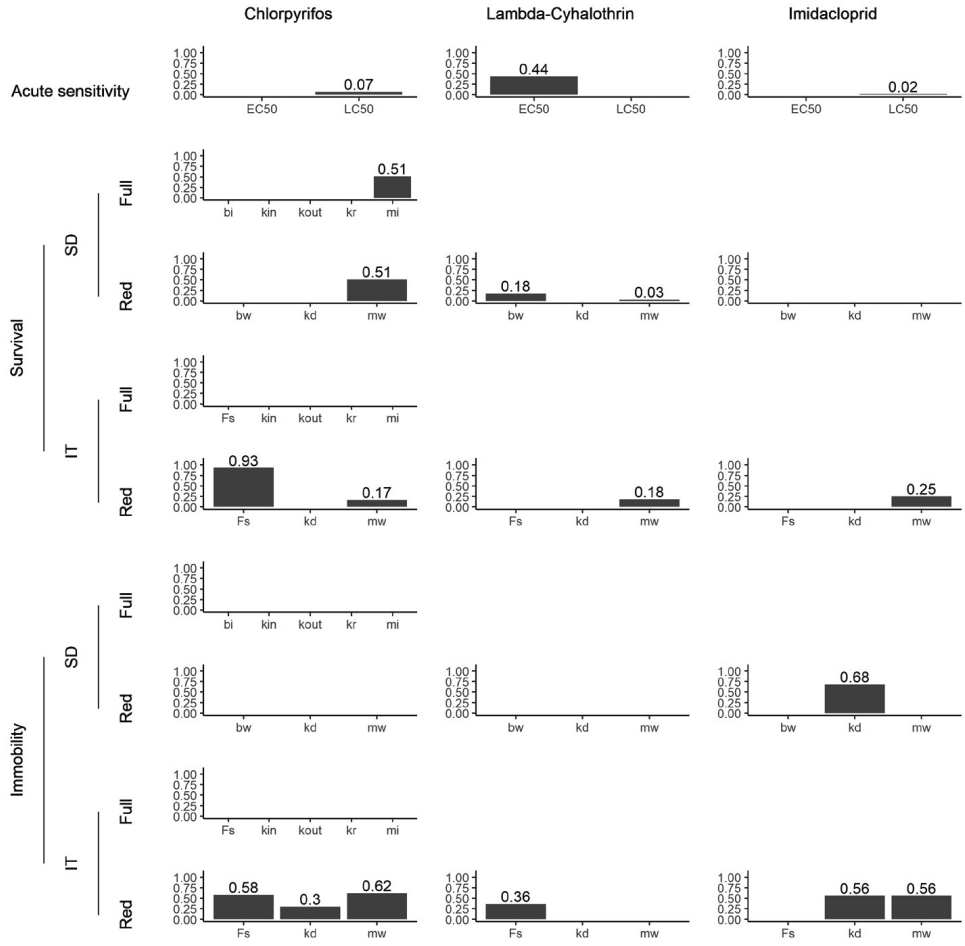
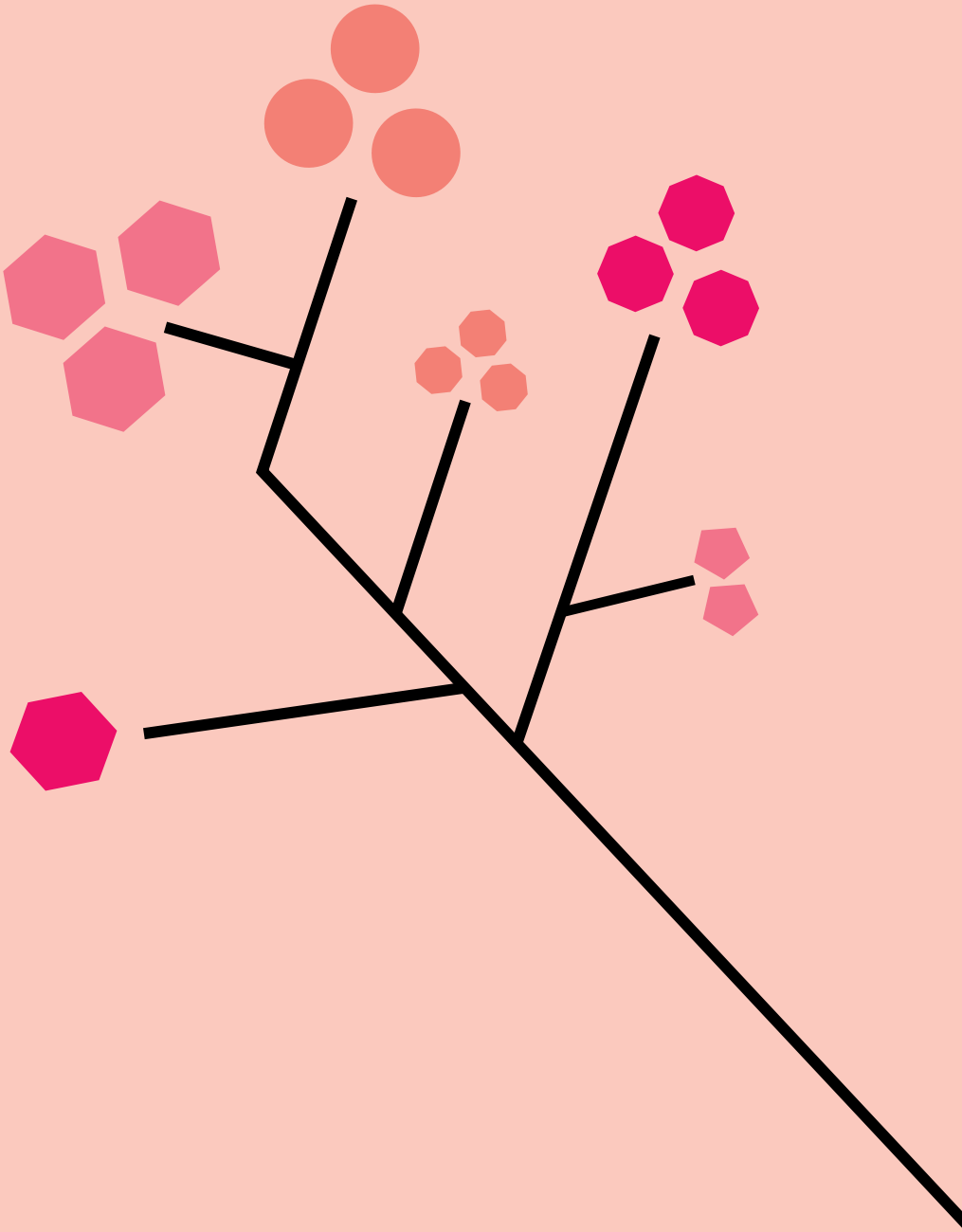


Figure S4.6. The prediction coefficient (P2) resulting from Leave-One-Out-Cross-Validation (LOOCV) performed for both endpoints (survival and immobility), all GUTS-models, and the three chemicals under study (chlorpyrifos, lambda-cyhalothrin, and imidacloprid). If no bar is shown, the P2 value was negative, indicating bad LOOCV results.



CHAPTER 5

Cross-species extrapolation of chemical sensitivity

Sanne van den Berg, Lorraine Maltby, Tom Sinclair, Ruoyu Liang,
Paul J. van den Brink

Abstract

The ecological risk assessment of chemicals largely depends on cross-species extrapolation approaches, since experimentally testing of all possible species-chemical combinations is impossible. This review provides an overview of currently existing cross-species extrapolation methodologies, and discusses i) how species sensitivity could be described, ii) which predictors might be useful for explaining differences in species sensitivity, and iii) which statistical considerations are important. We argue that risk assessment can benefit most from modelling approaches when species sensitivity is described based on effects that are ecologically relevant and robust. Additionally, specific attention should be paid to the heterogeneity of the training data, since this strongly influences the reliability of the resulting models. This includes heterogeneity in exposure duration, experimental variables, and the use of chemical or taxonomic grouping. Regarding which predictors are useful for explaining differences in species sensitivity, we review interspecies-correlation, relatedness-based, traits-based, and genomic-based extrapolation methods focusing on the amount of mechanistic information the predictors contain, the amount of input data the models require, and the extent to which the different methods provide protection for ecological entities. We use this review to develop an integrated framework, incorporating the strengths of each of the methods described. Finally, the discussion of statistical considerations reveals that regardless of the method used, statistically significant models can be found, although the usefulness, applicability, and understanding of these models and their outcomes varies considerably according to modelling choices made. We therefore recommend the publication of scientific code along with scientific studies to simultaneously clarify modelling choices, and enable continuation and elaboration on existing work. In general, we aim that this review can help steer the direction of empirical research by pointing out in which research fields data demands are most urgent, and make regulators more aware of where rules and regulations need to be put in place to make future cross-species extrapolation efforts successful.

5.1 Introduction

An ecosystem generally consists of a diverse species assemblage, and each of the species present in such assemblage has the potential to show a different sensitivity towards each of the many different chemical compounds that can be present in their environment (e.g. Hickey and Clements 1998, Biggs, Williams et al. 2007, Clements and Rohr 2009). Ecological risk assessment (ERA) is the process used to evaluate the impact of chemicals on species assemblages by seeking the threshold concentration below which ecosystem structure and functioning experience no adverse impacts (e.g. Suter 2016). At the first tier of this assessment, this threshold is defined by combining the results of single species toxicity tests with assessment factors (Brock, Arts et al. 2006). These assessment factors should reflect the uncertainty and variability related to the extrapolation from a laboratory system (short-term, high exposure, controlled environment, one species) to the natural environment (long-term, low exposure, variable environment, multiple species, and species interactions) (USEPA 2002). However, the assessment-factor approach remains generalized, since one threshold value is applicable to all assemblages within an ecosystem, irrespective of the variation in their species composition over space and time. This limits the specificity of the ERA. In contrast, existing higher tier approaches, such as mesocosm studies, do consider species assemblages rather than single species. However, performing multiple mesocosm experiments to account for seasonal and spatial variation would be too time and capital intensive (Van den Brink 2008). Predictive methodologies extrapolating existing toxicity data to untested organisms can help understand spatial-temporal variation in species sensitivity by predicting sensitivity values for a wide range of species (e.g. Malaj, Guénard et al. 2016, Raimondo and Barron 2019, Van den Berg, Baveco et al. 2019). However, although several predictive methods have been developed in the last decades, a clear overview of which extrapolation methodologies are currently available, along with a description of their considerations, assumptions, merits, and pitfalls, is still lacking.

Since the need to address spatial-temporal variation requires the sensitivity of a species assemblage to be calculated rather than the sensitivity of a single species, we focus this review on methods extrapolating the sensitivity of multiple species towards one chemical or mode of action (MOA), thereby excluding methodologies extrapolating sensitivity of one species to multiple chemicals (e.g. Quantitative-Structure-Activity Relationships (QSARs), Donkin 2009). Interspecies Correlation Estimation (ICE) is one of the earliest methods used to extrapolate toxicity data to untested species (Mayer and Ellersieck 1986). A software program to predict acute effects on aquatic and terrestrial species using ICE was developed in the 2000s (Asfaw, Ellersieck et al. 2003) and a web-based model is available as Web-ICE (Raimondo, Lilavois et al. 2015). The method has gained popularity for the derivation of water quality criteria (e.g. Dyer, Versteeg et al. 2008, Feng, Wu et al. 2013). In addition to their application to water quality, for example within the WFD (Water Framework Directive, European Commission 2000), cross-species extrapolation methods can also be utilised for assessing the risk from chemical exposure under REACH (Registration, Evaluation, and Authorization of

Chemicals, European Commission 2007), the Biocides Directive (European Commission 1998), the Plant Protection Products Directive (European Commission 1991), and any other directive that requires a chemical risk assessment.

To understand interspecific differences in species sensitivity towards chemical exposure, it is useful to divide sensitivity into two processes: toxicokinetics (TK) and toxicodynamics (TD) (EFSA Panel on Plant Protection Products and their Residues, Ockleford et al. 2018). TK processes describe the uptake, biotransformation and elimination of a chemical by a given organism, whilst TD processes are related to the damage, internal recovery and toxicity thresholds inside the organism after uptake of the chemical. The mechanistic basis of cross-species extrapolation is related to interspecific differences in TKTD processes. Interspecific differences in TKTD processes can be investigated by describing the combined effect of TK and TD processes simultaneously, or by using more specific predictors that split TK and TD into separate processes. In this review, we illustrate these processes in more detail, explain how they can be used as a more accurate description of species sensitivity, and clarify how different predictors can be used to describe different components of interspecific variation in sensitivity to chemical exposure.

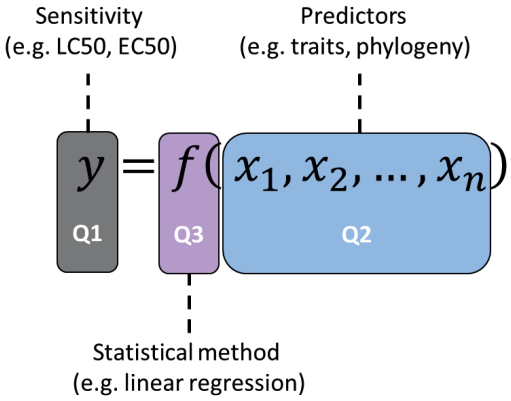


Figure 5.1. Schematic overview of the elements making up predictive models. Q1 indicates the elements covered by the first sub-question this review addresses, Q2 the second, and Q3 the third.

The main research question of this review is '*How can we extrapolate species sensitivity?*'. However, a direct answer to this question does not exist, and in order to understand and compare cross-species extrapolation methods, it is necessary to study the three elements that make up predictive models separately, namely: i) the dependent variable (y), ii) the independent variable(s) (x), and iii) the function used to determine the relationship between the independent variable(s) and the dependent variable (f, **Figure 5.1**). For the cross-species extrapolation methods reviewed here, the dependent variable is the sensitivity of an untested species to a chemical. Therefore, the first sub-question this review tries to answer is '*How can we describe*

species sensitivity?' (Q1). Although there is a proven distinction between true sensitivity and sensitivity as measured by short-term, laboratory experiments (Craig 2013), it remains unambiguous that true sensitivity can only be inferred from measured sensitivity. Therefore, to avoid ambiguity, we will continue to use the term sensitivity to refer to measured sensitivity, of which we are aware that it is a measure relative to the exact protocol under which it was determined. The second element making up predictive models is the independent variable(s), or in other words, the predictors required to explain species sensitivity. The second sub-question this review tries to answer is therefore '*Which independent variables are useful for explaining differences in species sensitivity?*' (Q2). Ultimately, the last element concerns the statistical considerations that are of importance when connecting the independent and dependent variables together, or in other words, an answer to the question '*Which statistical considerations are important when extrapolating species sensitivity?*' (Q3). Overall, we aim to provide an overview of the approaches available for each of the three elements mentioned, along with a description of the considerations and assumptions they make, and to provide guidance on how the optimal combination of these elements can be combined in an integrated framework.

5.2 How can we describe species sensitivity?

The first element concerns how sensitivity is described. This description is primarily dependent on choices made in the selection of the input data, since this limits the boundaries of the model. For example, if the input data exclusively contain data on mortality effects, the resulting model can only predict effects on mortality. Additionally, when comparing the performance of different models to determine which model is most suitable for answering a specific research question, it is important to consider whether data have been grouped or not (e.g. over chemicals or taxa).

5.2.1 Effects

Effects on mortality are most frequently incorporated into predictive models (**Table 5.1)**. This is primarily determined by data availability. More than 40% of all aquatic toxicity tests in the ECOTOX database (USEPA 2019) report effects on mortality, making it the most frequently studied effect on aquatic organisms of this database. However, this does not mean that mortality is the most important effect to consider, nor that these data cannot be exploited in a more optimal way, for instance, by means of TKTD models.

Effects other than mortality might be ecologically more relevant. Reproduction, for instance, is an indisputable element of population sustainability (see Gleason and Nacci 2001, for an example with fathead minnow, and see Segner 2011 for extensive background material). Thus, processes influencing reproductive success might be a better indicator of effects at higher levels of biological organization (e.g. offspring fitness, Hammers-Wirtz and Ratte 2000).

Table 5.1. Overview of modelling decisions made in the construction of interspecies correlation (IC), relatedness-based (RB), taxonomy-based (TB), and genomic-based (GB) models.

	Effects (Endpoint)	Exposure duration	Taxa	Transformation/ normalization	Unit of exposure conc.	Chemicals included per model	Grouping across taxa	Statistical method	Reference(s)
IC	Mortality (LC50), Immobilization (EC50)	48 to 96 h	Fish, algae, birds, mammals, aquatic invertebrates	$\log LC50$	µg/L	> 1 chemical or MOA	Species	Linear regression	(e.g. Dyer, Versteeg et al. 2006, Feng, Wu et al. 2013, Bejarano and Barron 2014, Brill, Belanger et al. 2016)
RB	Mortality (LC50), Immobilization (EC50)	48 to 96 h	Fish, aquatic invertebrates	$\log LC50$	µg/L	1 chemical	Species	Bayesian regression	(Craig 2013)
	Mortality (LC50)	96 h	Amphibians, fish, aquatic invertebrates	$\log LC50$	µmol/L	> 1 MOA	Species	Bilinear regression	(Guénard, von der Ohe et al. 2014)
	Mortality (LC50), Immobilization (EC50)	24 to 96 h	Aquatic invertebrates	$\log LC50$	µg/L	> 1 heavy metal	Species	Bilinear regression	(Malaj, Guénard et al. 2016)
	Population growth (EC50)	96 h	Algae	$\sqrt{LC50}$	µg/L	1 chemical	Species	Multivariate analysis	(Larras, Keck et al. 2014)
TB	Mortality (LC50), Immobilization (EC50)	24 to 96 h	Aquatic invertebrates	$\frac{\log LC50 - \mu}{\sigma}$	µg/L	1 MOA	Family	Linear regression	(Rubach, Baird et al. 2010)
	Mortality (LC50), Immobilization (EC50)	24 to 96 h	Aquatic invertebrates	$\log \frac{1}{LC50} * F^{a}$	mol/L	1 chemical	Genus, species	Genetic algorithm	(Ippolito, Todeschini et al. 2012)

	Mortality (LC50), Immobilization (EC50), Uptake (k_{in}), Elimination (k_{out})	48 h	Aquatic invertebrates	$\log LC50$	µg/L	1 chemical	Species	Linear regression	(Rubach, Baird et al. 2012)
	Mortality (LC50), Immobilization (EC50)	24 to 96 h	Aquatic invertebrates	$\frac{\log LC50 - \mu}{\sigma}$	µg/L	1 MOA	Order, family, genus	Linear regression	(Rico and Van den Brink 2015)
	Mortality (LC50)	24 to 96 h	Aquatic invertebrates	$\frac{\log LC50 - \mu}{\sigma}$	mol/L	1 MOA	Genus	Linear regression	(Van den Berg, Baveco et al. 2019)
GB	Mortality (LC50)	48 and 96 h	Amphibians, fish, aquatic invertebrates	$\log LC50$	µg/L	1 chemical	Species	Linear regression	(LaLone, Villeneuve et al. 2013)
	Mortality (LD50)	-	Birds	$\log LD50$	nmoles/kg	3 chemicals	Species	Linear regression	(Farmahin, Wu et al. 2012)

[a]Normalization factor was used to normalize the data according to exposure duration (Ippolito, Todeschini et al. 2012).

Energy allocation has been suggested as a means to link the various levels of biological organization together (Calow and Sibly 1990), since the energy available for reproduction and other functions depends on the availability of food sources and on the ability of an organism to exploit those (Amiard-Triquet 2009). Thus, effects on feeding behaviour and reproduction can directly be connected to effects at population level by means of energy allocation modelling (Calow and Sibly 1990), and might provide a closer approximation than when effects on mortality are used. More recently, this dynamic energy budget modelling has obtained renewed research interest under the acronym DEBkiss, promoting simple generic models of animal life history (Jager, Martin et al. 2013).

Besides incorporating more ecologically relevant measurement endpoints, it is also possible to extract more information from existing data by means of TKTD models. For instance, the General Unified Threshold model of Survival (GUTS) is a TKTD framework that has been developed to obtain more mechanistic understanding from mortality or immobilization data by dynamically describing the process of uptake, elimination, recovery, and survival (Jager, Albert et al. 2011). Since GUTS parameters provide a more accurate description of processes determining species sensitivity, additional mechanistic understanding of differences in species sensitivity can be obtained by comparing calibrated GUTS parameter values across species, instead of standard sensitivity endpoints (Rubach, Ashauer et al. 2011, Rubach, Baird et al. 2012). To be able to fit GUTS models, however, data on effects at multiple time points are required. Collection of these data is already obligatory in some standard test protocols (e.g. OECD 2019), in which case access to these data can easily be acquired by a commitment to publish the raw data of experiments along with summary statistics like LC50 values. Other standard test protocols, under which collection of this data is not yet mandatory, should be reviewed on whether their data requirements should be tightened to allow the fitting of such models.

5.2.2 Exposure duration

Typically, acute toxicity tests with an exposure duration between 24 and 96 hours are used for predictive modelling (**Table 5.1**). Again, this is primarily determined by data availability, since more than 50 percent of all aquatic toxicity test data available in the ECOTOX database concerns tests with an exposure duration of up to 96 hours (USEPA 2019). Although expanding the exposure duration range may be beneficial for obtaining an adequately-sized dataset, it potentially compromises the integrity of the model and should be avoided if possible. For instance, we are likely to find less effects after a 1-day exposure than after a 21-days exposure, because it takes time for a chemical to reach equilibrium between the exposure concentration and the concentration inside the organism. This difference is likely to become larger when the comparison concerns tests performed with different species, i.e. due to intraspecific differences in size and other traits influencing the uptake and elimination of the chemical (e.g. Wiberg-Larsen, Graeber et al. 2016). The exposure duration required to reach equilibrium is not only

species dependent, but also depends on the physical-chemical properties of the compound, as is well-known from QSAR modelling (Cherkasov, Muratov et al. 2014).

Besides running experiments long enough to ascertain that internal and external concentrations are in equilibrium, internal tissue concentrations could be reported together with external exposure concentration. Several studies have demonstrated that the internal chemical concentration describes toxic effects more closely than the external chemical concentration (Friant and Henry 1985, McCarty, Landrum et al. 2011). Focusing on internal chemical concentration would by-pass TK processes, since uptake and elimination processes are redundant when internal concentrations are known, and would enable us to compare differences in species sensitivity originating from internal processes only (TD). Alternatively, a TKTD model like GUTS could be employed, which results in toxicity measures that are independent of exposure time (Jager, Heugens et al. 2006).

5.2.3 Additional selection criteria

Imposing additional selection criteria on experimental conditions (e.g. pH, temperature, conductivity) can be useful for improving data homogeneity and hence data quality. Heavy metal toxicity, for example, has been reported to vary greatly according to the physicochemical characteristics of the exposed water (Pascoe, Evans et al. 1986, Gerhardt 1993). Hydroxide precipitations are, for instance, formed when the pH rises to alkaline conditions, reducing the concentration of free metal ions and reducing the heavy metal toxicity (Gadd and Griffiths 1977, Gerhardt 1993). Similarly, reduced heavy metal toxicity was observed in hard water due to the formation of insoluble compounds from soluble metal ions (McCarty, Henry et al. 1978). The biotic ligand model has been developed to examine the bioavailability of heavy metals under different exposure circumstances, and additionally explains how abiotic conditions influence the affinity of metals to accumulate on the surface of aquatic organisms (Erickson 2013). Similar models, normalization factors, or additional selection criteria, can be employed for other compound groups when necessary. Whether and which physicochemical properties should be taken into consideration when determining toxicity depends on the specific characteristics of the chemical group under study.

5.2.4 Units

A final, but equally important choice in the description of sensitivity data is the unit in which sensitivity is expressed. This is specifically important when comparing species sensitivity across chemicals, which is sometimes necessary when data availability is restricted (discussed in next section). Although $\mu g\ l^{-1}$ is still the most frequently used unit in aquatic toxicity tests (almost 50% of all aquatic tests available in the ECOTOX database, USEPA 2019, and see Table 5.1), it is not the most suitable one. It is frequently overlooked that chemical sensitivity is primarily related to molecular activities, and that the use of molar units makes molecule-to-molecule activity comparisons possible. For baseline toxicants exhibiting a non-polar narcosis

MOA, the concentration at which mortality occurs will be close to equivalent for all species when internal molar concentrations are used (Wezel and Opperhuizen 1995, Escher and Hermens 2002), reducing differences in species sensitivity to TK processes only. To overcome the problems of tests expressed in weight units, attaching an accurate molar mass database (e.g. EPIsuite, USEPA 2018) can help with converting mass units to molar units.

5.2.5 Grouping data, and its effects on explained variance

For modelling purposes, classifying chemicals according to their MOA is considered useful, because it provides an organizing scheme using an intermediate level of complexity between molecular mechanisms and physiological or organismal outcomes (Carriger, Martin et al. 2016). The rationale for using MOA classification for cross-species extrapolation is that these molecular mechanisms are conserved among biological entities (Escher and Hermens 2002). However, as in any grouping, using MOA as a grouping variable also introduces variation and errors. The assigned MOA may vary, for instance, between species or life stage depending on the availability of target sites (e.g. in the case of photosynthetic inhibitors, Nendza and Muller 2000), or between classification scheme used (see Kienzler, Barron et al. 2017 for differences in MOA classification according to the approach used). Therefore, MOA grouping only represents a suitable option when it is used with caution, for instance, by restricting the taxonomic range of the model to avoid interspecific variation in MOA, or when there is strong evidence that the MOA is applicable across the species in question (e.g. for baseline narcosis, for which there is strong evidence that the critical body residue for acute lethality in aquatic organisms has a very small range, van Wezel, de Vries et al. 1995).

Similar to using MOA to group across chemicals, higher taxonomic levels (e.g. family, order) can be used to group across taxa, and may also be useful for reducing data gaps. Grouping at higher taxonomic levels has the advantage of reducing bias due to extreme values and spurious data. However, potentially important differences in species sensitivity might be lost by summarizing the sensitivity of several species at, for example, family level (Buchwalter, Cain et al. 2008, Ippolito, Todeschini et al. 2012), and this trade-off should be carefully considered for the chemical-taxa combination under study.

Whether and how input data are grouped needs to be considered when comparing the performance (e.g. the adjusted R^2, or the cross-validation error) of different models. It is crucial to keep in mind that the variation associated with the grouping that goes into the model, is directly related to the variation related to the predictions that come out of the model (Schultz and Cronin 2003). Disregarding the variation in input values can result in an overly optimistic view on model performance. Similarly, when comparing the performance of different models, it is important to consider how much variation the model explains, since this largely depends on the number of chemicals considered in the model. For instance, the most complex model of Guénard and colleagues (2014) explained 80% of the variation in the sensitivity of 25 species towards five compounds, whilst a related model of Van den Berg et al. (Van den Berg, Baveco

et al. 2019, both models include AChE inhibition as MOA) explained only 41% of the variation in the sensitivity of 32 genera towards 33 compounds. This large difference in model performance can partially be explained by the fact that the five compounds of Guénard et al. included three MOAs, whilst the 33 compounds of Van den Berg et al. included only one MOA, thereby resulting in a large difference in the absolute amount of variation that each model explains.

5.3 Which independent variables are useful for explaining differences in species sensitivity?

We divide possible sensitivity predictors into four groups based on the type of mechanistic information that they contain: interspecies-correlation (IC), relatedness-based (RB), trait-based (TB), and genomic-based (GB). Here, we first give an overview of the general concept behind each sub-group (section 5.3.1), followed by a discussion of the merits and pitfalls associated with each of them (section 5.3.2), and close with a description on how the different predictor groups can be combined in an integrated framework (section 5.3.3).

5.3.1 Overview of methods

Interspecies-correlation (IC) models are log-linear least-squares regression of the acute toxicity (E/LC50) of chemicals measured in two species (e.g. Dyer, Versteeg et al. 2006, Raimondo, Mineau et al. 2007, Awkerman, Raimondo et al. 2008, Dyer, Versteeg et al. 2008, Awkerman, Raimondo et al. 2014). IC models aim at predicting the acute toxicity of a chemical to untested species (predicted species) using the known acute toxicity of this chemical to tested species (surrogate species). IC models have been used to predict chemical toxicity for algae (e.g. Brill, Belanger et al. 2016), aquatic invertebrates and vertebrates (e.g. Awkerman, Raimondo et al. 2014), terrestrial birds (e.g. Raimondo, Mineau et al. 2007) and mammals (e.g. Awkerman, Raimondo et al. 2009), and have proven to be protective for rare and endangered species (Willming, Lilavois et al. 2016). However, not all predictions made by this kind of models are reliable. Reliable prediction results are those that are derived from models that have a low mean square error, narrow confidence intervals, a high cross-validation success rate, a high R^2 value, and are predicting the sensitivity of closely related taxa (Raimondo, Mineau et al. 2007, Raimondo, Vivian et al. 2010, Raimondo and Barron 2019).

Relatedness-based (RB) models use the extent of evolutionary relatedness between organisms as a proxy for the similarity in their response to chemical stressors (e.g. Craig 2013, Guénard, von der Ohe et al. 2014, Malaj, Guénard et al. 2016). The underlying principle of these models is that closely related species exhibit high correlation of sensitivity to chemicals. This correlation can be used to make extrapolations from species whose sensitivity is known, to closely related untested species. The strength of this correlation decreases as the two species are more distantly related to the point where species that belong to the same phyla, but not to the same class, exhibit no correlation of sensitivity. Most RB models use taxonomy to predict

the sensitivity of untested species (e.g. Craig 2013), although other relatedness metrics, such as phylogenetics, have also been used (Guénard, von der Ohe et al. 2014, Malaj, Guénard et al. 2016, Table 5.1).

Trait-based (TB) models use physiological, morphological and ecological characteristics of a species to describe its sensitivity towards chemical stressors (e.g. Rubach, Baird et al. 2010). Currently existing trait databases (e.g. Usseglio-Polatera, Bournaud et al. 2000), primarily describe visible, external traits (e.g. size, shape). Therefore, TB models are most appropriate for describing TK related processes, e.g. by considering feeding mode or mode of respiration (Rubach, Baird et al. 2012, Van den Berg, Baveco et al. 2019). Other traits that could help describe internal TD processes (e.g. biotransformation, presence of target receptors) are available, but have so far only been described for a small number of species (see Table 2 Rubach, Ashauer et al. 2011 for an overview of the availability and linkage of potential toxicodynamic traits).

Genomic-based (GB) models use the relationship between gene expression and biological function as a way to determine the sensitivity of an organism towards specific chemical stressors (Snape, Maund et al. 2004, Fedorenkova, Vonk et al. 2010). GB models focus on gene and protein expression, integrating transcriptomics (identification of mRNA from actively transcribed genes), proteomics (identification of proteins in a biological sample), and metabolomics (identification of metabolites in a biological sample) into ecotoxicology (Pennie, Ahr et al. 2001). It is widely recognized that changes in gene expression have the potential to serve as early warning indicators for environmental effects and as useful biomarkers for chemical exposure (Pennie, Ahr et al. 2001, Poynton, Robinson et al. 2014), because they can be detected at low concentrations of chemicals and occur well before any morphological or reproductive effects become visible (e.g. Klaper and Thomas 2004). However, how effects found at a molecular level should be extrapolated to a higher biological level relevant to risk assessment is an area of active research, and adverse outcome pathways (AOPs) have been suggested as a suitable framework (Ankley, Bennett et al. 2010). An AOP is a conceptual construct of a sequence of events that starts with a molecular initiating event, spans multiple levels of biological organization, and ends with an adverse outcome on endpoints meaningful to risk assessment (e.g. survival, reproduction). We realize that the boundary between a phylogenetic RB approach and a GB approach can be vague. To avoid ambiguity, we consider an analysis of the sequence similarity in a molecular target a GB approach (because this confirms a deeper understanding of the toxicity process), whilst an analysis of the sequence similarity in the whole genome or in genetic markers frequently used in phylogenetic analysis (e.g. COI, 18S) is considered an RB approach (**Table 5.1**).

5.3.2 Comparison of methods

Mechanistic explanation

Raimondo and colleagues (2010) state that taxonomic relatedness is the underlying mechanistic explanation for IC models. However, IC models do not incorporate any phylogenetic- or taxonomic predictors, and only take taxonomic distance into account when screening for reliable prediction results (Raimondo and Barron 2019). Similarly, relatedness between chemicals can be considered the mechanistic explanation of IC models, since these models always include the response of species to multiple chemicals. Indeed, the fact that IC models work well when enough data are available, is likely due to the simultaneous explanation of the variation in sensitivity related to different chemicals and different species. Nevertheless, the lack of either taxonomic or physicochemical predictors raises the possibility of over-fitting the correlation model to the training data, resulting in inaccurate predictions when models are applied beyond the limits of the training data (Johnson and Omland 2004). In the case of IC models, any untested chemical lies outside the limits of the training data.

RB models use relatedness as the mechanistic explanation of sensitivity. Relatedness itself does not explain differences in sensitivity, but is used as a proxy for similarity in species response to chemicals (Craig 2013, Guénard, von der Ohe et al. 2014, Malaj, Guénard et al. 2016), since closely related taxa tend to exhibit similar sensitivity due to shared sensitivity-influencing traits (e.g. size and target receptor, Blomberg, Garland Jr et al. 2003). The shared distance from a common ancestor results in closely-related genetic patterns, which leads to a similar biochemistry and phenotype, and therefore, to a shared susceptibility to certain MOAs.

TB models incorporate mechanistic explanations of sensitivity arising from differences in phenotypic or ecological characteristics of species. One TB approach has, for instance, demonstrated that the uptake rate of chemicals can to a large extend be explained by the lipid content of an organism, whilst elimination rates are negatively correlated with the degree of sclerotization (Rubach, Baird et al. 2012). Depending on the taxonomic group under study, mechanistic hypotheses between traits and chemical susceptibility have been established to a greater or lesser extent. See, for instance, Table 1 in (Culp, Armanini et al. 2011) and the references therein, for examples of established trait-stressor relationships for algae, fish, aquatic plants, and invertebrates.

GB models have the potential to contain a comprehensive mechanistic explanation of sensitivity to chemical exposure. However, in contrast to TB models, GB models often describe complex biochemical pathways that are difficult to understand and to test experimentally (see McCarty and Munkittrick 1996 for an overview of the limitations of biomarkers for assessing population level effects). Even if a complete AOP is available, uncertainties in the quantification of one of the intermediate steps required to infer organism level effects from molecular target sequence similarity might prevent a model from performing well, i.e. have a large predictive power. This

is largely because these intermediate steps (e.g. related to transcriptomics, proteomics) heavily influence the eventual outcome of the molecular effect. LaLone and colleagues (2013) found, for example, that the correlation between empirical acute toxicity data and the percent similarity in the molecular target analysis is not very strong ($R^2 = 0.49$, p-value = 0.121). They argue that to fully understand chemical susceptibility it is necessary to further assess sequence and even structural information beyond the level of the primary or secondary protein structure (LaLone, Villeneuve et al. 2013).

Data demand

IC models only require data on toxicity (e.g. EC50, LC50), which can be obtained from public databases such as the ECOTOX Knowledgebase (USEPA 2019). However, the requirement that paired toxicity data (i.e. surrogate and predicted species) must be available for at least three chemicals in order to produce the correlation, restricts data availability (Raimondo, Jackson et al. 2010). Nevertheless, the latest IC models for aquatic animals contain more than 8500 toxicity values covering 316 species and 1499 chemicals (Raimondo, Lilavois et al. 2015). However, the taxonomic coverage of these models is restricted, with more than 60 percent of all the models available in WebICE extrapolating from one fish species to another (Raimondo, Lilavois et al. 2015), and of another 26 percent, either the surrogate or the predicted species is a fish.

As the predictive methods of RB models are based on relatedness, rather than on correlations of sensitivity to chemicals, data on toxicity must be complemented with data on relatedness. Taxonomic classifications for use in taxonomic RB models are readily available for any described species in publicly available databases (e.g. the taxonomy database from the National Center for Biotechnology Information, Federhen 2011, or the Integrated Taxonomic Information System, ITIS 2019). A phylogenetic RB model requires the genetic sequencing of a species, and coverage of phylogenies is currently still clade dependent. For instance, sequencing efforts in eukaryotic genomics are strongly biased towards multicellular organisms and their parasites (Del Campo, Sieracki et al. 2014), and large projects are available to sequence vertebrate genomes (e.g. the Genome 10K project, Koepfli, Paten et al. 2015). Genomic projects on algae and invertebrates remain limited, however, restricting the use of phylogeny-based RB models to data-rich clades such as fish. To ensure a good performance of RB models, however, a taxonomically (or phylogenetically) diverse toxicity dataset is required, because the correlation of sensitivity decreases with decreasing relatedness (Craig 2013).

The data demand of TB models depends on the traits to be included in the model, as well as the taxonomic group for which the model is constructed. For invertebrates, traits like size and mode of respiration (e.g. having gills or not) are readily available in literature, or can otherwise easily be recorded. Data on more specific traits, like lipid content or target site distribution, require more effort to measure, and are therefore less available in the literature (see Table 2 in Rubach, Ashauer et al. 2011). The study of Van den Berg and colleagues (2019) showed that when a

wide range of traits were included in the construction of invertebrate TB models, the modelling effort was primarily limited by a shortage of traits data (loss of 56% of the species for which toxicity data are available). However, only one trait database was used in their study (Usseglio-Polatera, Bournaud et al. 2000), whilst more trait databases are available for invertebrates (Poff, Olden et al. 2006, Schäfer, Kefford et al. 2011, Hébert, Beisner et al. 2016). For fish, a wide range of traits are available distributed over several trait databases (Froese and Pauly 2000, Lamouroux, Poff et al. 2002, Frimpong and Angermeier 2009), and covering a large part of the taxonomic diversity of fish. For algae we are aware of two traits databases currently available (Reynolds, Huszar et al. 2002, Lange, Townsend et al. 2016), but have to acknowledge that they are likely to have the lowest taxonomic coverage out of the three standard organism groups discussed here (invertebrates, fish, algae), due to the large biodiversity of this group. Besides data on traits, TB models require data on taxonomy to match the traits with the toxicity data. Access to taxonomic data has already been described under RB models.

GB models are the most data demanding, because they require validated AOPs. Currently, 274 AOPs have been described in the AOP wiki in total covering 521 stressors, although the OECD status of the majority of them remains 'under development' (https://aopwiki.org/, accessed on the 25th of January 2019), and taxonomic coverage of these models remains limited. However, powerful advances in genome sequencing technology, informatics, automation, and artificial intelligence are assisting researchers in understanding species differences to a more detailed level (Lewin, Robinson et al. 2018), and can be expected to lead to a significant increase in the development of AOPs. Promising new techniques, e.g. in vitro cell-lines (Eisner, Doering et al. 2019) or enzymatic markers (Arini, Mittal et al. 2017), are being developed and carry the potential to replace currently used in-vivo concentration-response curves with in-vitro concentration-response curves (see, for instance, Figure 3 in Zhang, Xia et al. 2018). However, these methods are time-, and cost-consuming, and are frequently incomparable due to inconsistent bioinformatic methods for data filtering, concentration-response modelling and quantitative characterization of genes and pathways (Zhang, Xia et al. 2018).

Protection of ecological entities

The main objective of all cross-species extrapolation methods is to get an accurate view on the variation in species sensitivity that exists in the real world. Indeed, all methods presented in this review attempt to add realism to risk assessment by filling in data gaps. However, the methods studied in this review vary in two important ways: i) in the way they can consider real species assemblages, and ii) in the way that they can be used to extrapolate effects to higher levels of biological organization (e.g. population, community or ecosystem level). Therefore, the four methods differ in the way they provide protection for ecological entities, i.e. species, populations, communities, and ecosystems.

Researchers have known for a long time that real species assemblages vary through time (Murphy 1978) and space (Vannote, Minshall et al. 1980). Although we will likely never be

able to understand this variation in its entirety, we can reduce uncertainty in ERA by predicting the sensitivity of representative species assemblages. RB and TB methods have this potential, since both methods can predict the sensitivity of species that have never undergone toxicity testing before, providing predictor values of the species whose sensitivity you want to predict are available. This contrasts with IC models, which require sufficient toxicity data to be available for other chemicals for the taxon whose sensitivity we want to predict (section 5.3.1), and then still might be overfitted to the training data. GB models require, at least, to have the part of the genome sequenced that is associated with the key molecular initiating events, so that sequence similarity can be calculated (LaLone, Villeneuve et al. 2013). Therefore, IC and GB models are only able to predict the sensitivity of commonly tested species.

All four methods have the potential to be used for the construction of species sensitivity distributions (SSDs), a statistical tool considered more protective of ecological entities than single measurements of sensitivity, since they allow only a defined fraction of species present in a species assemblage to be affected (Kooijman 1987). Again, due to the restrictions in the underlying data, IC and GB models assume standard species assemblages in their SSDs, whilst RB and TB models can also be applied to representative species assemblages. RB approaches have as advantage over TB approaches that data on relatedness is usually more abundant than data on traits, allowing sensitivity to be predicted for a wider range of species. Although, in the contrary, RB approaches have a weaker mechanistic explanation, ignoring the fact that taxonomically diverse species can share sensitivity-related traits. For this reason, RB models can be used to develop spatially-defined protection criteria, whereas TB models can extrapolate found relationships towards assemblages with the same trait profile, but with a different taxonomic composition (Van den Brink, Alexander et al. 2011). GB approaches have recently been used for the retrospective risk assessment of community-level effects towards ammonia and nitrogen using field-based SSDs (Yang, Zhang et al. 2017). However, there are many uncertainties in using retrospective risk assessment approaches, for instance, due to the inability to disentangle effects caused by the stressor of interest from all the other stressors (either natural or anthropogenic) that might be present at the site under study. For this reason, we do not consider retrospective risk assessment studies in our review.

Although SSDs are considered more representative of real species assemblages than when only an alga, an invertebrate, and a fish are evaluated, they still do not consider indirect effects of chemical exposure, i.e. effects on food availability, predation, competitive interactions or feedback mechanisms. Indeed, all studies described in this review only consider direct effects of chemical exposure on organism sensitivity. However, certain methods are better able than others to extrapolate effects to higher levels of organization. For instance, TB models permit the derivation of hypotheses on what might happen to specific functional groups, whilst RB can only do this if functions are clearly restricted to taxonomic or phylogenetic groups. Imagine, for example, that predators are more sensitive to a certain chemical than herbivores due to a difference in assimilation efficiency (a relationship found in Hendriks, van der Linde et al.

2001). It is well known from literature that functional traits like feeding guild are not strongly conserved across taxonomy (e.g. see Table 1 in Poteat, Jacobus et al. 2015 for the distribution of feeding guilds over the orders Ephemeroptera, Plecoptera, and Trichoptera). Therefore, RB approaches will fail to extrapolate the effect of this relationship to the community level, whilst TB approaches will be able to do so. Additionally, hypotheses derived from TB models can directly link into stochastic ecosystem models (e.g. De Laender, Morselli et al. 2015). Such models are able to extrapolate effects found for specific functional groups to the community level, incorporating factors like species interactions and functional redundancy (Rosenfeld 2002). For GB approaches, examples exist of how to extrapolate direct effects to population level effects. For instance, De Coen and Janssen (2003) have found a strong relationship ($0.88 < r^2 < 0.99$) between the cellular energy allocation biomarker response to several chemicals and population level effects of *Daphnia magna*. However, studies extrapolating effects found on a single species to community level effects remain absent. For IC models, no examples of extrapolations to higher biological levels exist, besides the use of assessment factors.

5.3.3 Combined approaches to predicting sensitivity

Since all the methods discussed in this review have their own strengths and weaknesses, our main concern is not identifying which method results in models with the highest explanatory power, but rather in understanding how the methods can be incorporated into one integrated framework. Indeed, all studies discussed in this review (**Table 5.1**) have demonstrated the ability to predict differences in species sensitivity to a certain extent, although there was not one method that consistently outperformed the others, and all of them seemed restricted in the maximum amount of variation in species sensitivity they could explain. However, studies which combined predictors from multiple mechanistic explanations observed an increased model performance compared to when predictors belonging to only one mechanistic explanation were included. For example, Larras et al. (2014) and Buchwalter et al. (2008) both found that combining TB and RB methods (trophic preference with phylogenetic signal, and body weight with taxonomic family, respectively) explained more variation than either method alone. These findings have found consistent support in further studies (e.g. Ippolito, Todeschini et al. 2012, Poteat, Jacobus et al. 2015).

How combining predictors belonging to different predictor groups leads to better models can be explained by the fact that each of the predictor groups explains a different part of the sensitivity processes as understood under the TKTD framework (**Figure 5.2**). Studies describing species differences in TK parameters (e.g. Buchwalter, Cain et al. 2008, Rubach, Baird et al. 2012) found that traits like mode of respiration and body size are good predictors of uptake rates, whilst elimination rates have a very strong phylogenetic signal. We are unaware of any studies that have explored the relationships between GB predictors and TD parameters, but since TD parameters describe processes related to toxicity thresholds inside the organism, the presence, absence, and distribution of chemical receptors are likely to be strong predictors

for describing differences in the TD part of species sensitivity (e.g. as found in Larras, Keck et al. 2014). So, we can hypothesize that TB approaches are good in explaining the TK part of differences in species sensitivity, whilst GB approaches have high explanatory power for predicting cross-species differences in TD processes. Additionally, RB approaches have the potential to represent aspects of both TK and TD, because relatedness acts as a proxy for the likelihood of sharing a niche and therefore traits (TK), but also for sharing similar biochemical processes (TD). For these reasons, we hypothesize that the best performing models can be found by combining the different methods in one integrated framework (**Figure 5.2**), although next we will discuss the constraints that apply to this framework.

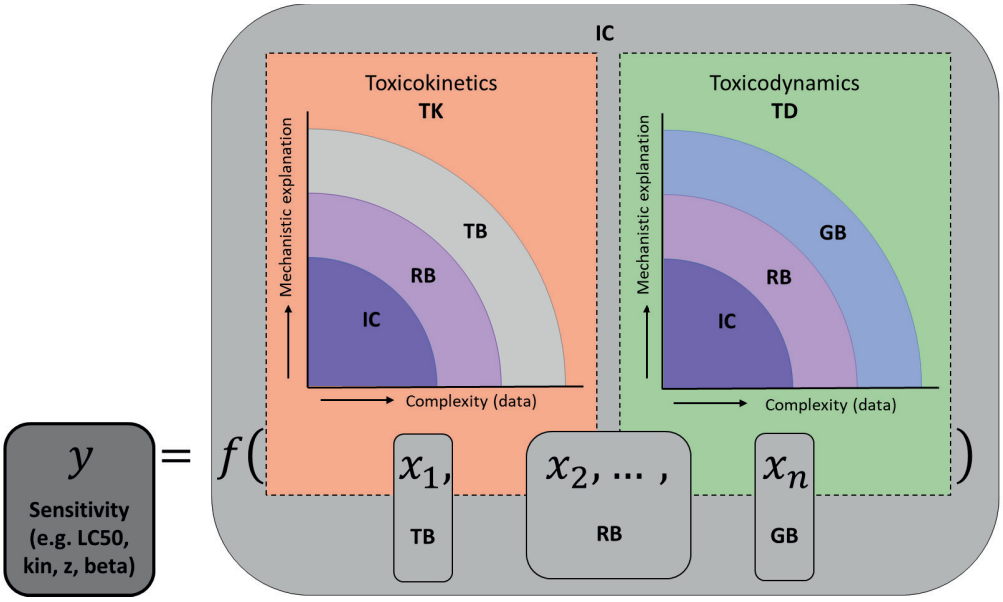


Figure 5.2. An abstract visualization of the integrated framework suggested to combine the different modelling approaches (IC, RB, TB and GB) discussed in this review. The different layers (IC, RB, TB, GB) of the TK and TD processes can be regarded as the steps of a tiered approach, increasing in complexity and mechanistic explanation.

When combining predictors from different categories, it is important to set constraints on the exact part of the sensitivity process they are going to explain, because otherwise the predictors can negatively influence each other through collinear relationships, as explained in section 5.4. For the reasons mentioned in the previous paragraph, we hypothesize that the best performing models can be found by combining sensitivity-related, morphological traits (TB) with the sequence similarity of the molecular target(s) of the chemical or MOA under study (GB). RB predictors can potentially be added to the model to represent sensitivity related processes that are still unknown. Alternatively, a separate RB approach can be used to distinguish which taxa are sensitive and tolerant to a specific chemical or MOA. This information can be help ease the search for the molecular target(s) describing differences in species sensitivity, since it must lie

in the part of the genome that is different between these taxa. We think that IC approaches have limited applicability for cross-species extrapolation of chemical sensitivity when data are restricted due to their lack in mechanistic information, but they can of course be used when data availability is sufficient. The different layers (IC, RB, TB, GB) of the TK and TD processes as illustrated in **Figure 5.2** can also be regarded as the steps of a tiered approach, increasing in complexity and mechanistic explanation.

5.4 Which statistical considerations are important when extrapolating species sensitivity?

The final feature of predictive models that this review discusses, is the statistical considerations that are important when extrapolating species sensitivity. After all, most modelers are aware that a major part of the modelling outcome is determined by choices made along the modelling process. These choices range from the selection of input data (section 5.2), to the method selected for (preliminary) variable selection. Here, we want to discuss modelling considerations that have so far not been discussed in this review, but are main determinants for the modelling output.

The first consideration is the omission of data points. Modelling studies often depend on a subset of data available in literature or databases, and, as mentioned in section 5.2, model performance is largely dependent on this sub-setting of the input data. Therefore, it is crucial that data are only omitted or included under clear and well-documented circumstances. Data should never be omitted without explanation, as this can lead to the suspicion that outliers were merely removed to improve the model.

The second consideration is the use of confounded predictors. If two predictors are highly collinear, they contribute the same information twice, thus confounding the statistical association and making it more difficult to deduce a mechanistic interpretation (Dormann, Elith et al. 2013). Therefore, preliminary variable selection is an important process. Van den Berg et al (2019) assessed the optimal collinearity threshold for trait predictors, and found an increase in cross-validation error with an increasing collinearity threshold. In general, a collinearity of maximum 70% is allowed, and is found sufficient to keep collinearity under control (e.g. Dormann, Elith et al. 2013). Research performed on a GB based approach studied the influence of different preliminary variable selection methods on model performance (Mannheimer, Duval et al. 2019). They found that the variable selection method only had marginal effects on Spearman correlations between predicted and measured values, and that as long as the signal to noise ratio is high, the dominant effect will be captured regardless of the preliminary variable selection method. This is to a large extent true for big datasets containing many collinear predictors, which is often the case for GB approaches. For smaller datasets, however, preliminary variable selection methods can have a severe impact on the modelling results. Therefore, predictors should be collected deliberately avoiding collinearity, and with clear underlying hypotheses.

The third consideration is that any descriptor value, measured or calculated, can potentially contain errors. Molecular descriptors, for instance, may vary depending on the conformation of molecules and on the software used (Benfenati, Piclin et al. 2001, Schultz and Cronin 2003). Therefore, the more descriptor values included in the model, the larger the chance of incorporating errors. Extrapolating the variation associated with descriptors is a field not yet satisfactorily explored, but crucial if modelling approaches ever want to take a more dominant place in the risk assessment process (e.g. by means of Bayesian approaches, Wintle, McCarthy et al. 2003). For this to be possible, though, accessibility to raw data is necessary. Proper registration and transparency of test methods used and results generated will help making data-mining approaches more feasible, especially if raw data are organized according to clear standards. Guidelines and standards have been developed for ecotoxicity data (e.g. Kase, Korkaric et al. 2016, Moermond, Kase et al. 2016, Society of Environmental Toxicology and Chemistry 2019), but also for gene expression data the minimum quantity and quality of information required to interpret and verify study results has been defined (Brazma, Hingamp et al. 2001).

The fourth and final consideration concerns overfitting in general. Biological processes consist of complex dynamic interactions in a multi-dimensional system, and non-linear methods have the ability to capture these complex interactions between variables (e.g. Ladroue, Guo et al. 2009). However, in a multi-dimensional system these methods tend to incorporate noise leading to overfitting. Alternatively, linear methods are more robust to overfitting, although at the cost of potentially missing important non-linear interactions (Mannheimer, Duval et al. 2019). Whether a linear or non-linear method is more suitable depends on the number of predictors available, and on the degree of mechanistic information contained within these predictors. Regardless, additional measures can be taken to ensure overfitting is avoided. The use of the adjusted R^2 as model selection criterion should, for instance, be avoided, although this rule is still regularly broken (e.g. Rubach, Baird et al. 2010, Rubach, Baird et al. 2012, Rico and Van den Brink 2015). This criterion focuses entirely on maximizing fit and completely disregards model complexity, therefore often resulting in models overfitted to the training data. Information criteria that consider both fit and complexity (e.g. Akaike's Information Criterion) are better suited for selecting a model (Johnson and Omland 2004), and are therefore recommended.

Regardless of the exact choices made on the considerations discussed in this section, it is likely that statistically significant models will be found. However, the outcome of these models does to a large extend depends on the choices made. For this reason, communication of choices made during the modelling process is just as crucial for understanding the modelling outcomes, as are the modelling outcomes themselves. Striving for reproducible research is one way to force modelling choices to be communicated, since being able to recreate the whole process will enable external reviewers to re-run all the steps made. Reproducible research has as additional advantage that methods that have been implemented once, do not require implementation a

second, third, or even fourth time. In this way, we can spend our efforts on using and elaborating on existing work.

5.5 Concluding remarks

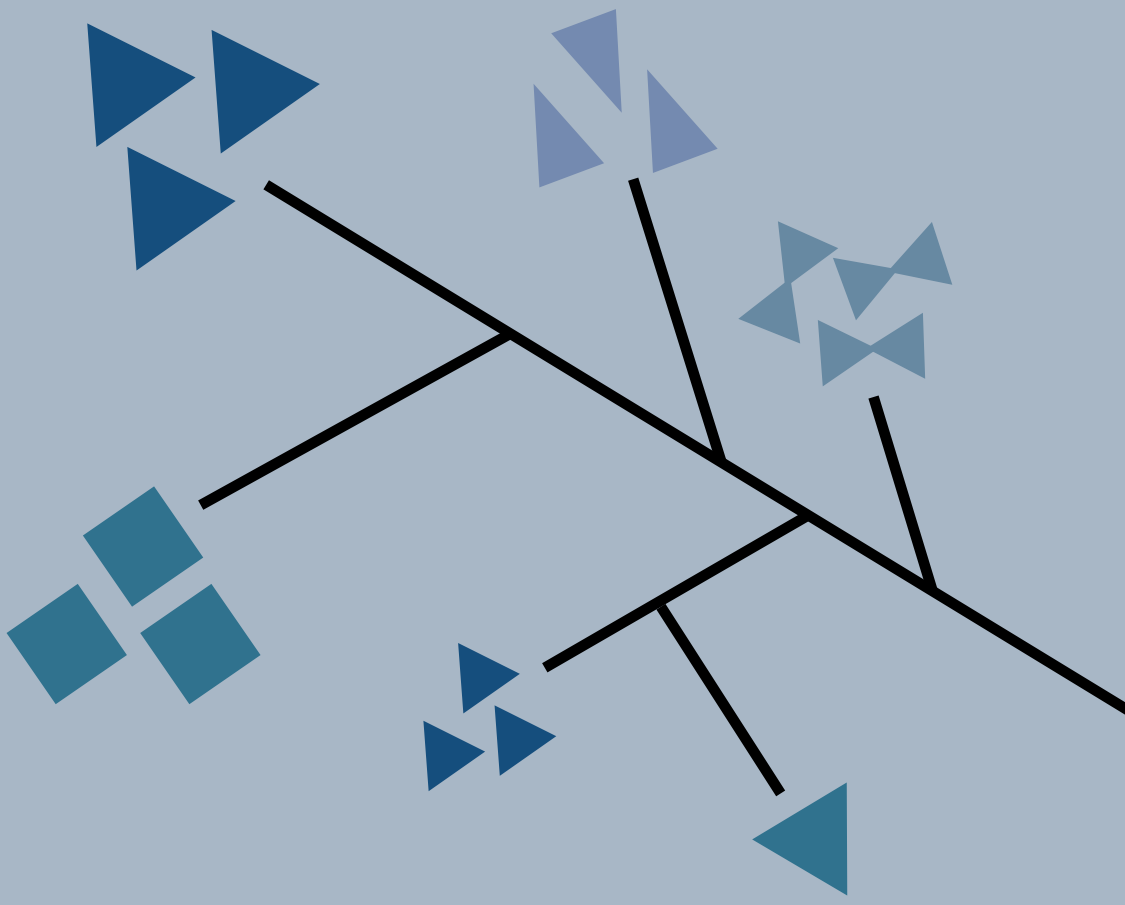
This review provides an overview of the methodologies currently available for extrapolating species sensitivity towards chemical stressors. However, there is not one straight-forward answer to the question '*How can we extrapolate species sensitivity?*'. Indeed, the answer to this question depends on the answers to the sub-questions addressed in this review: i) how can we describe species sensitivity, ii) which independent variables are useful for explaining differences in species sensitivity, and iii) which statistical considerations are important when extrapolating species sensitivity?

Regarding the first question, we show that ERA can primarily benefit from modelling approaches by describing species sensitivity on effects that are ecologically relevant and sufficiently robust such that the data can be used to accurately represent species sensitivity. However, attention should be paid to data heterogeneity, since this strongly influences the reliability of the resulting models. Additionally, the importance of the unit used to describe species sensitivity was discussed, which is primarily important when sensitivity is compared across chemicals, for instance, when data is grouped according to MOA. Ideally, concentrations should be described using molarities, since chemical sensitivity is primarily related to molecular activities. Finally, when deciding on which model is most suitable to answer a specific research question, we should keep in mind that model performance is a function of the number of chemicals and/or organisms that the model covers.

Regarding the independent variables that are useful for explaining differences in species sensitivity, we find that none of the methods discussed in this review result in the best model performance when considered alone. When sufficient toxicity data are available, and the MOA of the chemical is not very specific, IC models are likely to work (e.g. for baseline toxicants with a strong phylogenetic signal). However, as toxicity data for the same chemical is required for the tested and predicted species, IC methods are limited to species frequently used in laboratory testing. Extrapolating to other species therefore requires mechanistic approaches to construct trustworthy models. In that case, a combination of predictors originating from multiple approaches is likely to achieve optimal model performance, since all predictors explain a unique, complementary part of differences in species sensitivity (**Figure 5.2**). For these reasons, we suggest an integrated framework (**Figure 5.2**), combining predictors describing important traits determining the uptake and elimination of chemicals (e.g. size, respiration, exoskeleton-thickness), with the amount of sequence similarity in molecular targets, and relatedness predictors. This integrated framework can also be considered a tiered approach, where moving up a tier equals moving up in level of complexity and mechanistic understanding of the sensitivity process.

The final question has perhaps the most straight-forward answer, since regardless of the method selected, significant models can be found. It is, therefore, important that modelling is done in a reproducible way, and that modelling decisions are clearly communicated along with modelling results. To optimise reproducibility, we advise the publication of well-documented scientific code along with scientific studies, as is also in accordance with the good modelling practise as advised by EFSA (2014). This will not only clarify modelling choices, but will also help avoid re-implementing methods that have been implemented before, so that we can spend our efforts on continuing and elaborating on existing work.

In general, this review clarifies which considerations and assumptions are made during modelling efforts concerned with the cross-species extrapolation of chemical sensitivity. We think it can help steer the direction of empirical research, by pointing out in which research fields data demands are most urgent. For example, the applicability of RB and TB methods could be improved by refining which species have to be tested to enable accurate prediction to more natural species assemblages. For instance, focusing on still unknown taxonomic- or trait profiles. In contrast, GB models would benefit most from further research into AOPs and completing the necessary genomic tests for organism groups capable of being covered by GB methods. By specifying and standardising the data requirements for the functioning of these methods, it would be possible to integrate cross-species extrapolation into regulatory environments. This would offer opportunities for refining risk assessments, including spatial and temporal consideration of sensitivity, and provide methods for reducing animal testing and the cost associated with them.



CHAPTER 6

Afterthoughts

6.1 The myth of the most sensitive species

For over 30 years, researchers have been aware that there is not a single species or a specific group of species which is always the most sensitive one (all the time, everywhere, and towards every compound). This has been coined the 'myth of the most sensitive species' (Cairns 1986). Nevertheless, current risk assessment approaches still often rely on a standard set of test species, since well-developed experimental protocols are required to enable reproducibility of test results.

Chapter 2 provides new proof that relying on a standard set of test species results in an unrealistic representation of the actual sensitivity of invertebrates, and therefore provides evidence that current risk assessment approaches require rethinking. We demonstrate that there is not one genus, family, or class which is sensitive to all chemicals. Indeed, our results show that the mode of action of a chemical largely determines which taxa are most sensitive towards that chemical. We found, for instance, that chemicals classified as ion-, osmoregulatory-, or circulatory- impairing were most toxic towards bivalves, whilst acetylcholinesterase (AChE) inhibiting chemicals were most toxic towards arthropods (**chapter 2**). Similarly, when looking at *Daphnia* spp., the aquatic species that is most frequently incorporated into the aquatic risk assessment of chemicals, we found that although it is relatively sensitive towards chemicals belonging to the mode of action ester narcosis, it is one of the most tolerant species towards nicotinic acetylcholine receptor agonism (**Figure 6.1**). The same is true for other species commonly used in risk assessment, like *Chironomus* spp. and *Gammarus* spp. (**Figure 6.1**). For this reason, we conclude from **chapter 2** that depending on the same set of standard test species can result in either an overestimation or an underestimation of sensitivity, depending on the mode of action of the chemical.

6.2 Sensitivity is a complex problem

Species sensitivity is known to vary across life stage (e.g. Van der Lee, Kraak et al. 2020), gender (McClellan-Green, Romano et al. 2007), size (e.g. Poteat and Buchwalter 2014), and season (Van den Brink, Van Smeden et al. 2016), to name a few. This thesis adds to that list, variation in sensitivity according to the mode of action of the chemical (**chapter 2**), the exposure duration (**chapter 4**), and the endpoint under consideration (**chapter 4 & 5**). Consider, for instance, the change in the ranking of invertebrates after short- or long-term exposure, or after considering lethal- or sub-lethal effects (**chapter 4**). This long, and yet still incomplete, list of factors influencing species sensitivity makes it clear that sensitivity is a complex attribute.

With complexity in understanding sensitivity, we also face considerable challenges in reliably measuring it. Ecotoxicologists generally rely on simple measures to express the toxicity of chemicals to organisms, such as the no-observed-effect concentration (NOEC) and the exposure concentration associated with x% effect (ECx). However, these statistical summaries are merely a snapshot description of the toxicity process, and are largely influenced by the factors

mentioned before. True sensitivity is rather the response of a dynamic system (an individual) to a chemical stressor (Jager 2011). For this reason, the use of dynamic, mechanistic effect models has gained popularity in recent years (Jager, Heugens et al. 2006). In **chapter 4** we explored the use of these more accurate descriptions of species sensitivity, and have demonstrated that these parameters indeed allow us to understand underlying processes in species sensitivity better.

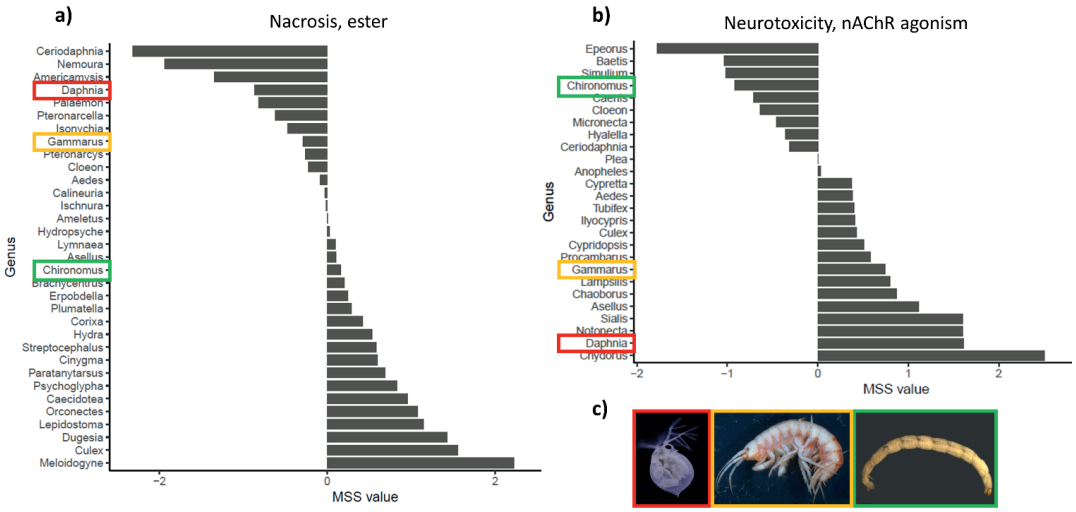


Figure 6.1. Illustration of how the position of three commonly tested species (Daphnia spp., Gammarus spp., and Chironomus spp., **c)** varies in sensitivity rankings made for different modes of action: **a)** ester, narcosis, and **b)** nicotinic acetylcholine receptor agonism, neurotoxicity. Pictures taken by Christophe Brochard (*Daphnia*) and Ton van Haaren (*Chironomus* and *Gammarus*).

Nevertheless, if we already struggle to grasp the sensitivity process at an individual level, then how can we start extrapolating to higher levels of biological organization (e.g. community or ecosystem level)? Predictions of the real impact of chemicals on an aquatic system are uncertain due to the complexity of the biological system and all the aspects influencing the system under study. However, just because something is uncertain does not imply that the right action is to do nothing until we are certain. Indeed, uncertainty can cut two ways: reality can turn out better or worse than predicted. Concerning direct effects, species can turn out more sensitive in the field than measured in the lab, for instance, due to effects of seasonality (i.e. temperature, Macaulay, Buchwalter et al. 2020), the presence of multiple stressors (Cedergreen 2014), and the historical presence of stressors (even if they are now absent, Ashauer, O'Connor et al. 2017). Species can also turn out less sensitive in the field than measured in the lab, for instance, due to adaption to the harsh conditions present in their natural environment (Räsänen, Laurila et al. 2003), or the constant exposure regimes used in the lab tests (Reinert, Giddings et al. 2002). Concerning indirect effects, impacts can turn out worse than expected due to a cascading effect

at the population or community level effect. Consider, for instance, the loss of keystone species or ecosystem engineers, for which any loss of function will result in a disproportional effect on ecosystem functions (Baert, De Laender et al. 2017). Alternatively, impacts can turn out lower than expected due to a high recoverability of the species (Gabsi and Preuss 2014), or due to the fact that other species take over the ecosystem function of the species that goes extinct (i.e. functional redundancy, Rosenfeld 2002). These indirect effects are currently not yet addressed with the cross-species extrapolation methods discussed in this thesis. However, being able to predict the sensitivity of natural species assemblages opens up the possibility of incorporating real species assemblages into community models such as (De Laender, Morselli et al. 2015).

6.3 Traits alone are insufficient in describing species sensitivity

In **chapter 2** we seem to hit a boundary on the maximum possible performance of trait-based models (figure S1 and S2 in the supporting information of chapter 2). Since previous studies have found that both traits and relatedness explained a unique part of sensitivity (Poteat, Jacobus et al. 2015, Pilière, Verberk et al. 2016), we decided to enhance the models developed in **chapter 2** with advanced taxonomic predictors (**chapter 3**). We indeed find that combining traits with taxonomic information results in models with increased predictive power (**Table 3.2**), and that regardless of whether the mode of action is strongly rooted in phylogeny, taxonomic predictors can be used to describe differences in species sensitivity. We find, for instance, that there is no specific taxonomic group more sensitive or tolerant towards narcotic compounds, whilst sensitivity towards AChE inhibiting compounds has a strong phylogenetic signal. In **chapter 4**, we find that taxonomic distance is a strong explainer of differences in elimination rates of imidacloprid for different species.

Another main finding of **chapter 2** was that a large part of the available toxicity data remains unused due to insufficient traits data. Around 70% of the species for which we had toxicity data available were discarded because no or incomplete traits data were available. Additionally, we sometimes struggled that traits were not quantitative enough. Consider, for instance, that the traits included in the best models found in **chapter 3** were conserved among taxonomy, leading to a bias when the model was used to predict the sensitivity of real species assemblages. From this we can conclude that we have currently reached a dead-end on using existing data only. Therefore, if we want to increase the predictive power of trait-based approaches, we must dedicate ourselves to collecting more traits data. Although it is possible that the lack of traits data gets resolved due to general research interest (e.g. Gallagher, Falster et al. 2020), it is more likely to be executed efficiently when the measurement of a restricted set of traits (with established relationships with sensitivity) becomes mandatory under standardized test guidelines as developed by organizations like the OECD (Organisation for Economic Co-operation and Development, e.g. OECD 2019).

Since traits on the one hand are insufficient in describing species sensitivity, and on the other hand pose one of the main obstacles in model construction, **chapter 5** describes and compares a more diverse range of possible sensitivity predictors. We suggest an integrated approach, combining predictors from multiple categories (e.g. traits, relatedness, genetic markers, **Figure 5.2**), and demonstrate that when specific predictors are used to explain specific parts of the sensitivity process we will be able to construct better models.

6.4 Sensitivity towards chemical stressors is spatially variable

In **chapter 3** the value of being able to predict the sensitivity of species never tested before becomes apparent. By applying developed models to species assemblage data, we demonstrate that aquatic ecosystems located in central Europe contain fewer sensitive species than those situated in the south. We found this was the case for both narcotic and AChE inhibiting compounds. However, zooming-in to a smaller spatial scale of one region, the UK in this case, revealed contradicting patterns in species sensitivity depending on the mode of action of the chemical. So, although entire regions can be considered relatively tolerant, there might still be certain river reaches with a large percentage of sensitive species. At both spatial scales, however, patterns of endemic biodiversity were found to explain the presence of sensitivity hotspots.

Since sensitivity towards chemical stressors is spatially variable, and the spatial pattern depends on the stressor under study, patterns across wide spatial scales can easily be compared with other studies to reveal regions where multiple stressors might be causing an effect simultaneously (**Figure 6.2**). This could be useful in the consideration of multiple stressors and interactions between multiple stressors can easily be incorporated, for instance, by using simple additive models (Cedergreen 2014, Rider, Dinse et al. 2018).

Assuming we reconcile conceptual differences in species sensitivity (**section 6.2**), there remain limiting factors in terms of data availability and collection quality across regions, especially in the developing world (e.g. Aus der Beek, Weber et al. 2016). We know the least about areas where pollution might present the largest challenge for ecosystems as well as the regulatory policy makers. A possible solution to estimate the potential chemical impact at these sites is to use species distribution models to predict assemblage composition at sites where no biomonitoring efforts are in place (e.g. as in He, Bradley et al. 2015). In this way, abiotic conditions can be used to determine which species are likely to occur at these sites. Subsequently predicting the sensitivity of these predicted assemblages would then allow us, for the first time, to map sensitivity globally at a fine resolution.

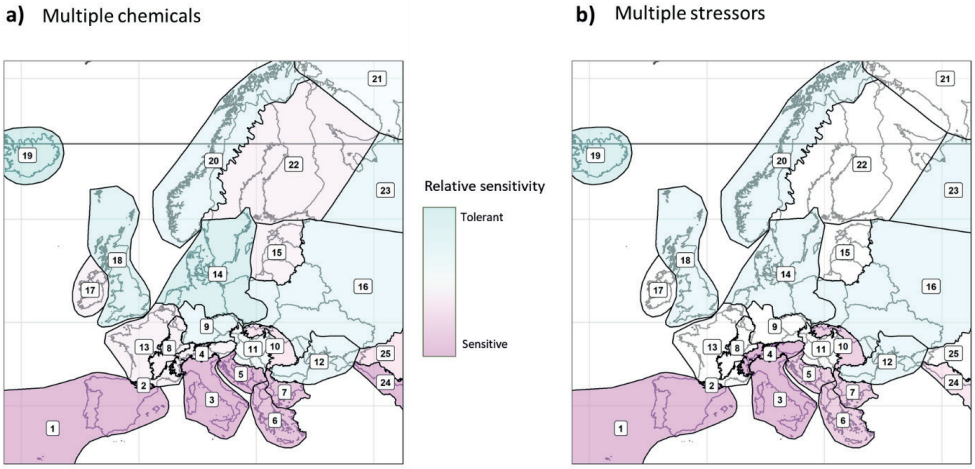


Figure 6.2. The relative sensitivity of aquatic invertebrates towards the mixture toxicity of **a)** AChE inhibiting and narcotic compounds, and **b)** climate change and narcotic compounds using simple additive models. Data on sensitivity towards climate change is derived from (Hershkovitz, Dahm et al. 2015).

6.5 It's all relative

In this thesis, we describe and discuss numerous times that results often depend on (modelling) decisions. In the supporting information of **chapter 2**, for example, we provide an extensive review and analysis on the effect of modelling decisions on modelling outcomes. However, even after this careful consideration of multiple factors influencing modelling decisions, model biases revealed themselves when applying the models to real data. For instance, we came to learn the existence of a taxonomic bias when we applied the models developed in **chapter 2** to existing species assemblages in **chapter 3**. This taxonomic bias was caused by a discrepancy in the taxonomic diversity and evenness between the data used to train the model, and the data to which the model was applied. This example is a clear illustration of how large the implications are of the well-known fact of any model: what goes in, must come out.

In principle, models are little more than a transformation of the data that was used to build the model. Therefore, all models need to be evaluated on their validation, applicability domain, and relevance (**Figure 6.3a**). The integrated approach we suggest in **chapter 5** has a large overlap with Quantitative-Structure-Activity-Relationship (QSAR) approaches. Consider, for instance, that instead of chemical structures, structure is resembled by species traits (i.e. describing the structural characteristics of species), and that instead of measures of chemical activity, activity is resembled by (the expression of) validated genetic markers (i.e. describing the presence/absence of chemical receptors). Due to this large resemblance, the validity, applicability, and relevance prerequisites developed for QSARs (Worth 2010) can easily be amended to cover cross-species extrapolation methods.

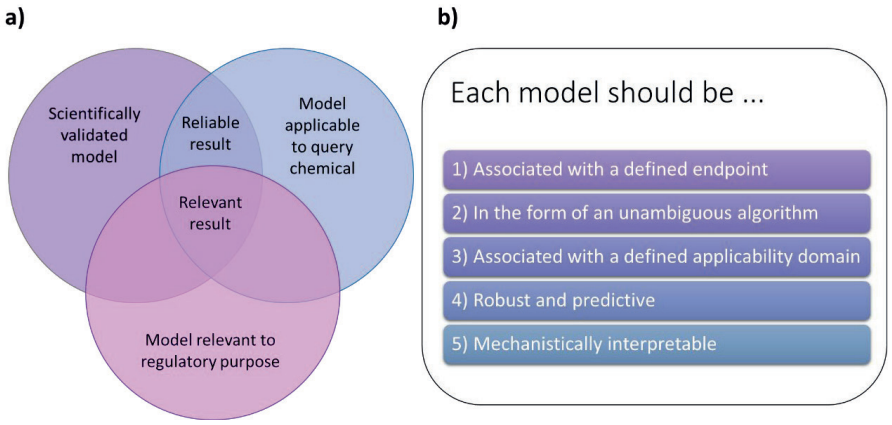


Figure 6.3. a) The overlap of considerations regarding the validity, applicability, and relevance needed to demonstrate model accuracy (adjusted from Worth 2010), and **b)** the five validation principles established by the OECD.

To facilitate the consideration of a cross-species extrapolation model for regulatory purposes, models have to demonstrate their validity (also see Augusiak, Van den Brink et al. 2014). Five validation principles have been established by the OECD, originally developed for QSAR models, that can be amended to include cross-species extrapolation methods (OECD 2004, figure 6.3b). First, a model should be associated with a defined endpoint. This refers to any biological effect that can be measured, as long as it is transparent. In **chapter 5** we describe the most relevant endpoints for consideration in risk assessment. Second, a model should be expressed in the form of an unambiguous algorithm. Although I would not call the algorithm (R pipeline) developed under **chapter 2 and 3** unambiguous, the main complexity in these cross-species extrapolation methods lies in the connection and preliminary processing of multiple databases. Eventual model construction consists most of the time out of mere multiple linear regressions (**Table 5.1**). The third validation principle describes that each model should have a clearly defined applicability domain. In the case of cross-species extrapolation, often grouping on either chemical or taxonomic groups are used to enable the construction of models with a restricted data availability (**Table 5.1**). The models developed in **chapter 2 and 3**, for example, have been constructed on the mode of action of the chemicals for which empirical toxicity data was available. Therefore, these models can only be applied to chemicals with a similar mode of action. Similarly, the models have been constructed for specific taxonomic groups, and the applicability of the models is thereby restricted to these taxa. The last two validation principles, stating that the models should be 4) robust and predictive and 5) mechanistically interpretable, are with the understanding of this thesis, highly correlated with each other. **Chapter 4** demonstrates that even for very small datasets, if the mechanisms determining species sensitivity are well-described by available predictors (e.g. traits), model performance will be good.

Besides that the model should be valid, it should also be applicable to the chemical under query. As mentioned in the previous section, models are associated with a defined application domain. Therefore, any uncertainty in whether the model is applicable to the query chemical is directly associated with the uncertainty of the classification method used to assign the mode of action, or any other grouping that was used for the training of the model.

Finally, the model endpoint should be relevant for the regulatory purpose. In the case that the model predicts directly the regulatory endpoint (e.g. LC50 values), the relevance is self-evident. However, the new generation cross-species extrapolation models focus on predicting lower-level mechanistic endpoints (e.g. uptake rates (**chapter 4**), presence of molecular sites (**chapter 5**)), and therefore require an additional extrapolation to the endpoint of interest. This is more thoroughly discussed in **chapter 5**.

6.6 Cross-species extrapolation for environmental risk assessment

The past years, progress has been made regarding the incorporation of so-called non-testing methods for use in risk assessment. Take, for instance, the more frequent incorporation of QSARs and read-across for the registration of new chemicals under REACH (the Registration, Evaluation, Authorization and Restriction of Chemicals, European Commission 2007). For an overview of case studies demonstrating the successful incorporation of non-testing methods into ERA, multiple review studies are available (OECD 2006, Worth and Patlewicz 2007). The cross-species extrapolation methods covered by this thesis would be a new addition to the list of non-testing methods, but how do we envision that cross-species extrapolation can be used in ERA?

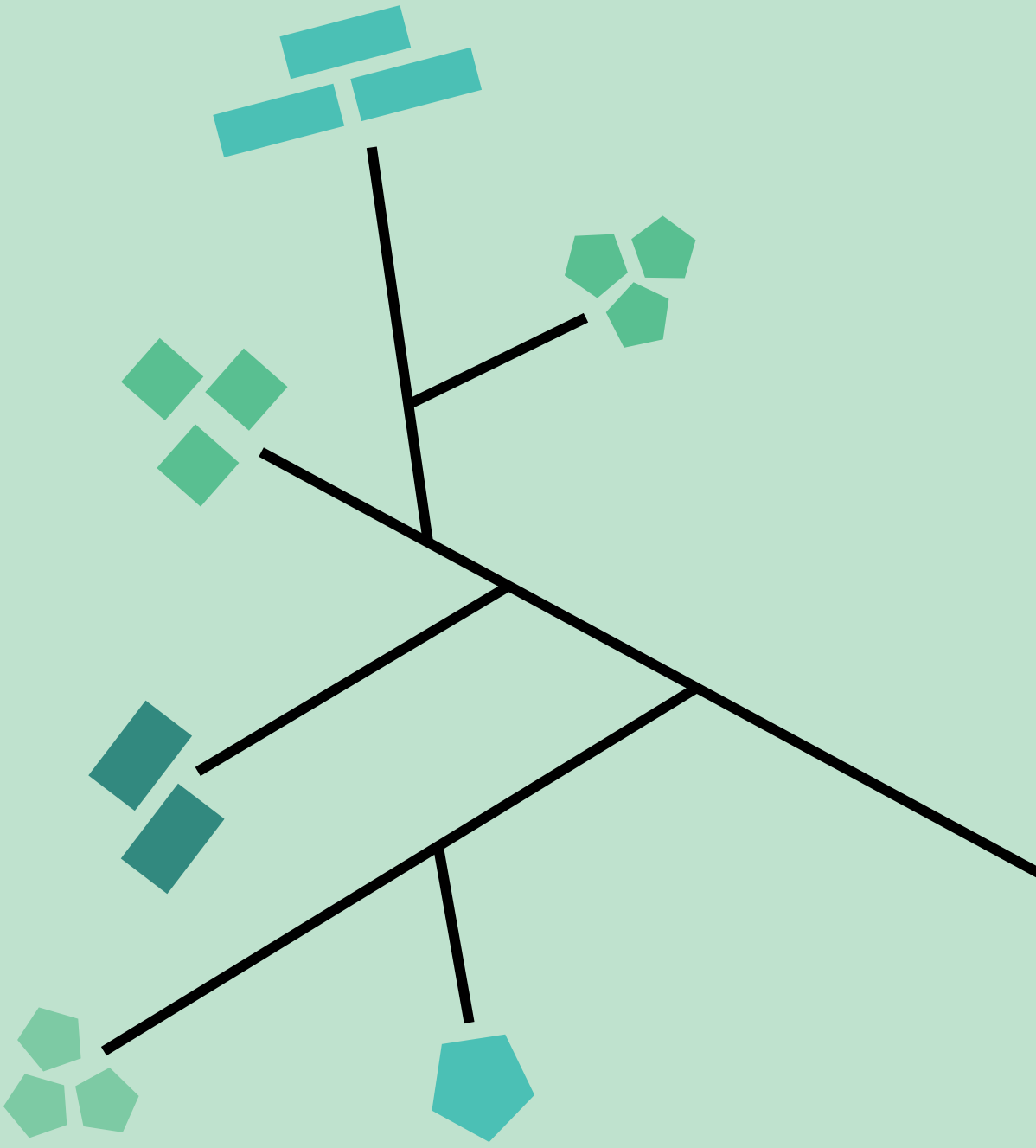
In principle, cross-species extrapolation can be used in ERA in three ways. First, it can be used to support priority setting procedures to narrow down further assessment work. The **first chapter** of this thesis provides sensitivity rankings that serve exactly this purpose (also see **Figure 6.1**). If the mode of action of the query chemical is known, we can use these sensitivity rankings (based on experimental data available on other chemicals with the same mode of action) to determine which taxa are most sensitive, and then focus further assessment work on these taxa. To avoid only considering taxa already frequently tested in the lab, we can extend the diversity of species diversity by applying the developed models to a wide diversity of untested species. This could point out species not normally considered for toxicity testing, but with an apparently high susceptibility towards the query chemical. **Chapter 5** additionally provides a workflow of how relatedness-based approaches can be used to narrow-down the taxonomic scope of further assessment work, even when no mechanistic predictors of sensitivity are known.

Second, cross-species extrapolation can be used to supplement the use of experimental data in weight-of-evidence approaches. Using cross-species extrapolation can strengthen weight-of-evidence approaches performed under lower tiers, or reduce the magnitude of standard

assessment factors. The power of cross-species extrapolation lies in this case not only in the filling in of data gaps, but more importantly, in the provision of mechanistic information. **Chapter 4** demonstrates that the availability of predictors (e.g. traits) that describe the mechanisms determining species sensitivity well is crucial for the construction of predictive models when data availability is low.

Finally, cross-species extrapolation can be used to replace or completely substitute the need for experimental data. This is of course the envisioned objective of all non-testing methods: completely replace animal testing. However, the way cross-species extrapolation methods will be applied in practice is likely to depend on the restrictions set by the regulatory framework (some being more conservative than others) and the specific context (availability of other information, possible consequences of an incorrect prediction). In regard of the first two applications mentioned (support priority setting and supplement experimental data), the use of cross-species extrapolation methods is more indirect and is more likely to be not decisive in the final assessment. However, when we discuss the replacement of experimental data with modelled data, the risk assessment will heavily rely on the performance of the models, and therefore will require properly validated and applicable models (Worth 2010).

In general, this thesis contributes to the development of a transparent ERA approach that is based on mechanistic understanding and combines experimental data with modelling approaches to develop smart, cost-effective, prospective tools. A well-developed ERA will protect species currently not directly covered by the aquatic testing procedures. As science is evolving, it is important to ensure that ERA is continuously updated with the latest technology and the newest scientific knowledge. It would be best to unify guidelines and regulation across the whole globe. Future steps can be taken to include metabolites, mixture toxicity, and focus on developing approaches at a landscape level further. This calls for an interdisciplinary approach to improve our scientific understanding and to communicate findings with all stakeholders involved in ERA.



References

Amiard-Triquet, C. (2009). "Behavioral Disturbances: The Missing Link between Sub-Organismal and Supra-Organismal Responses to Stress? Prospects Based on Aquatic Research." Human and Ecological Risk Assessment: An International Journal **15**(1): 87-110.

Ankley, G. T., R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, J. A. Serrrano, J. E. Tietge and D. L. Villeneuve (2010). "Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment." Environmental Toxicology and Chemistry **29**(3): 730-741.

Arini, A., K. Mittal, P. Dornbos, J. Head, J. Rutkiewicz and N. Basu (2017). "A cell-free testing platform to screen chemicals of potential neurotoxic concern across twenty vertebrate species." Environmental Toxicology and Chemistry **36**(11): 3081-3090.

Asfaw, A., M. R. Ellersieck and F. L. Mayer (2003). Interspecies Correlation Estimations (ICE) for Acute Toxicity to Aquatic Organisms and Wildlife: II. User Manual and Software, US Environmental Protection Agency, Office of Research and Development.

Ashauer, R., A. Agatz, C. Albert, V. Ducrot, N. Galic, J. Hendriks, T. Jager, A. Kretschmann, I. O'Connor, M. N. Rubach, A.-M. Nyman, W. Schmitt, J. Stadnicka, P. J. van den Brink and T. G. Preuss (2011). "Toxicokinetic-toxicodynamic modeling of quantal and graded sublethal endpoints: A brief discussion of concepts." Environmental Toxicology and Chemistry **30**(11): 2519-2524.

Ashauer, R., A. B. Boxall and C. D. Brown (2007). "Modeling combined effects of pulsed exposure to carbaryl and chlorpyrifos on Gammarus pulex." Environmental science & technology **41**(15): 5535-5541.

Ashauer, R., A. B. Boxall and C. D. Brown (2007). "New ecotoxicological model to simulate survival of aquatic invertebrates after exposure to fluctuating and sequential pulses of pesticides." Environmental science & technology **41**(4): 1480-1486.

Ashauer, R., A. Hintermeister, I. Caravatti, A. Kretschmann and B. I. Escher (2010). "Toxicokinetic and toxicodynamic modeling explains carry-over toxicity from exposure to diazinon by slow organism recovery." Environmental Science & Technology **44**(10): 3963-3971.

Ashauer, R. and T. Jager (2018). "Physiological modes of action across species and toxicants: the key to predictive ecotoxicology." Environ Sci Process Impacts **20**(1): 48-57.

Ashauer, R., I. O'Connor and B. I. Escher (2017). "Toxic Mixtures in Time—The Sequence Makes the Poison." Environmental Science & Technology **51**(5): 3084-3092.

Ashauer, R., I. O'Connor, A. Hintermeister and B. I. Escher (2015). "Death dilemma and organism recovery in ecotoxicology." Environmental science & technology **49**(16): 10136-10146.

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

Aus der Beek, T., F.-A. Weber, A. Bergmann, S. Hickmann, I. Ebert, A. Hein and A. Küster (2016). "Pharmaceuticals in the environment—Global occurrences and perspectives." Environmental Toxicology and Chemistry **35**(4): 823-835.

Awkerman, J. A., S. Raimondo and M. G. Barron (2008). "Development of species sensitivity distributions for wildlife using interspecies toxicity correlation models." Environmental science & technology **42**(9): 3447-3452.

Awkerman, J. A., S. Raimondo and M. G. Barron (2009). "Estimation of Wildlife Hazard Levels Using Interspecies Correlation Models and Standard Laboratory Rodent Toxicity Data." Journal of Toxicology and Environmental Health, Part A **72**(24): 1604-1609.

Awkerman, J. A., S. Raimondo, C. R. Jackson and M. G. Barron (2014). "Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models." Environmental Toxicology and Chemistry **33**(3): 688-695.

Bacon, S. J., A. Aebi, P. Calanca and S. Bacher (2014). "Quarantine arthropod invasions in Europe: the role of climate, hosts and propagule pressure." Diversity and distributions **20**(1): 84-94.

Baert, J. M., F. De Laender and C. R. Janssen (2017). "The Consequences of Nonrandomness in Species-Sensitivity in Relation to Functional Traits for Ecosystem-Level Effects of Chemicals." Environ Sci Technol **51**(12): 7228-7235.

Baird, D. J., C. J. Baker, R. B. Brua, M. Hajibabaei, K. McNicol, T. J. Pascoe and D. de Zwart (2011). "Toward a knowledge infrastructure for traits-based ecological risk assessment." Integrated environmental assessment and management **7**(2): 209-215.

Baird, D. J., M. N. Rubach and P. J. Van den Brink (2008). "Trait-based ecological risk assessment (TERA): The new frontier?" Integrated Environmental Assessment and Management **4**(1): 2-3.

Baird, D. J. and P. J. Van den Brink (2007). "Using biological traits to predict species sensitivity to toxic substances." Ecotoxicol Environ Saf **67**(2): 296-301.

Barron, M. G., C. R. Lilavois and T. M. Martin (2015). "MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development." Aquat Toxicol **161**: 102-107.

Bass, C., R. A. Carvalho, L. Oliphant, A. M. Puinean, L. M. Field, R. Nauen, M. S. Williamson, G. Moores and K. Gorman (2011). "Overexpression of a cytochrome P450 monooxygenase, CYP6ER1, is associated with resistance to imidacloprid in the brown planthopper, Nilaparvata lugens." Insect Molecular Biology **20**(6): 763-773.

Beadle, K., K. S. Singh, B. J. Troczka, E. Randall, M. Zaworra, C. T. Zimmer, A. Hayward, R. Reid, L. Kor, M. Kohler, B. Buer, D. R. Nelson, M. S. Williamson, T. G. E. Davies, L. M. Field, R. Nauen and C. Bass (2019). "Genomic insights into neonicotinoid sensitivity in the solitary bee Osmia bicornis." PLOS Genetics **15**(2): e1007903.

Bejarano, A. C. and M. G. Barron (2014). "Development and practical application of petroleum and dispersant interspecies correlation models for aquatic species." Environmental science & technology **48**(8): 4564-4572.

Belanger, S., M. Barron, P. Craig, S. Dyer, M. Galay-Burgos, M. Hamer, S. Marshall, L. Posthuma, S. Raimondo and P. Whitehouse (2017). "Future needs and recommendations in the development of species sensitivity distributions: Estimating toxicity thresholds for aquatic ecological communities and assessing impacts of chemical exposures." Integrated Environmental Assessment and Management **13**(4): 664-674.

Bellwood, D., P. Wainwright, C. Fulton and A. Hoey (2002). "Assembly rules and functional groups at global biogeographical scales." Functional Ecology **16**(5): 557-562.

Benfenati, E., N. Piclin, A. Roncaglioni and M. Varì (2001). "Factors influencing predictive models for toxicology." SAR and QSAR in environmental research **12**(6): 593-603.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2009). "GenBank." Nucleic Acids Res **37**(Database issue): D26-31.

Bernard, R., M. Heiser, A. Hochkirch and T. Schmitt (2011). "Genetic homogeneity of the Sedgling Nehalennia speciosa (Odonata: Coenagrionidae) indicates a single Würm glacial refugium and trans-Palaearctic postglacial expansion." Journal of Zoological Systematics and Evolutionary Research **49**(4): 292-297.

Biggs, J., P. Williams, M. Whitfield, P. Nicolet, C. Brown, J. Hollis, D. Arnold and T. Pepper (2007). "The freshwater biota of British agricultural landscapes and their sensitivity to pesticides." Agriculture, Ecosystems & Environment **122**(2): 137-148.

Blomberg, S. P., T. Garland Jr and A. R. Ives (2003). "Testing for phylogenetic signal in comparative data: behavioral traits are more labile." Evolution **57**(4): 717-745.

Bradbury, S. P. (1994). "Predicting modes of toxic action from chemical structure: an overview." SAR and QSAR in Environmental Research **2**(1-2): 89-104.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball and H. C. Causton (2001). "Minimum information about a microarray experiment (MIAME)—toward standards for microarray data." Nature genetics **29**(4): 365.

Brill, J. L., S. E. Belanger, J. G. Chaney, S. D. Dyer, S. Raimondo, M. G. Barron and C. A. Pittinger (2016). "Development of algal interspecies correlation estimation models for chemical hazard assessment." Environmental toxicology and chemistry **35**(9): 2368-2378.

Britten, R. J. (2002). "Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels." Proceedings of the National Academy of Sciences **99**(21): 13633-13635.

Brock, T. C., G. H. Arts, L. Maltby and P. J. Van den Brink (2006). "Aquatic risks of pesticides, ecological protection goals, and common aims in European Union legislation." Integrated Environmental Assessment and Management: An International Journal **2**(4): e20-e46.

Buchwalter, D., J. Jenkins and L. Curtis (2002). "Respiratory strategy is a major determinant of [3H] water and [14C] chlorpyrifos uptake in aquatic insects." Canadian Journal of Fisheries and Aquatic Sciences **59**(8): 1315-1322.

Buchwalter, D. B., D. J. Cain, W. H. Clements and S. N. Luoma (2007). "Using Biodynamic Models to Reconcile Differences Between Laboratory Toxicity Tests and Field Biomonitoring with Aquatic Insects." Environmental Science & Technology **41**(13): 4821-4828.

Buchwalter, D. B., D. J. Cain, C. A. Martin, L. Xie, S. N. Luoma and T. Garland (2008). "Aquatic insect ecophysiological traits reveal phylogenetically based differences in dissolved cadmium susceptibility." Proceedings of the National Academy of Sciences **105**(24): 8321-8326.

Buchwalter, D. B. and S. N. Luoma (2005). "Differences in Dissolved Cadmium and Zinc Uptake among Stream Insects: Mechanistic Explanations." Environmental Science & Technology **39**(2): 498-504.

Cairns, J. J. (1986). "The Myth of the Most Sensitive SpeciesMultispecies testing can provide valuable evidence for protecting the environment." BioScience **36**(10): 670-672.

Calow, P. and R. Sibly (1990). "A physiological basis of population processes: ecotoxicological implications." Functional Ecology **4**(3): 283-288.

Carriger, J. F., T. M. Martin and M. G. Barron (2016). "A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors." Aquatic Toxicology **180**: 11-24.

Cedergreen, N. (2014). "Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology." PloS one **9**(5).

Chamberlain, S. and E. Szöcs (2013). taxize - taxonomic search and retrieval in R., F1000Research.

Chapman, P. M., A. Fairbrother and D. Brown (1998). "A critical evaluation of safety (uncertainty) factors for ecological risk assessment." Environmental Toxicology and Chemistry **17**(1): 99-108.

Cherkasov, A., E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard and A. Tropsha (2014). "QSAR Modeling: Where Have You Been? Where Are You Going To?" Journal of Medicinal Chemistry **57**(12): 4977-5010.

Chevenet, F., S. Dolédec and D. Chessel (1994). "A fuzzy coding approach for the analysis of long-term ecological data." Freshwater biology **31**(3): 295-309.

Clark, J. S. (2005). "Why environmental scientists are becoming Bayesians." Ecology letters **8**(1): 2-14.

Clements, W. H. and J. R. Rohr (2009). "Community responses to contaminants: using basic ecological principles to predict ecotoxicological effects." Environmental Toxicology and Chemistry: An International Journal **28**(9): 1789-1800.

Cole, J. F. H., C. J. Everett, W. Gentle, J. A. Ashwell and U. Goggin (2000). Azoxystrobin: an outdoor pond microcosm study (summarised and cited in EFSA 2009). ZENECA Agrochemicals. Jealott's Hill International Research Centre. .

Craig, P. S. (2013). "Exploring novel ways of using species sensitivity distributions to establish PNECs for industrial chemicals: Final report to Project Steering Group 3 April 2013."

Culp, J. M., D. G. Armanini, M. J. Dunbar, J. M. Orlofske, N. L. Poff, A. I. Pollard, A. G. Yates and G. C. Hose (2011). "Incorporating traits in aquatic biomonitoring to enhance causal diagnosis and prediction." Integr Environ Assess Manag **7**(2): 187-197.

Davy-Bowker, J., R. T. Clarke, T. Corbin, H. Vincent, J. Pretty, A. Hawczak, J. Murphy and I. Jones (2008). "River invertebrate classification tool."

De Coen, W. M. and C. R. Janssen (2003). "The missing biomarker link: Relationships between effects on the cellular energy allocation biomarker of toxicant-stressed Daphnia magna and corresponding population characteristics." Environmental Toxicology and Chemistry: An International Journal **22**(7): 1632-1641.

De Laender, F., M. Morselli, H. Baveco, P. J. Van den Brink and A. Di Guardo (2015). "Theoretically exploring direct and indirect chemical effects across ecological and exposure scenarios using mechanistic fate and effects modelling." Environment International **74**: 181-190.

Deharveng, L., H. Dalens, H. Drugmand, J. C. Simon-Benito, M. M. Da Gama, P. Sousa, C. Gers and A. Bedos (2000). "Endemism mapping and biodiversity conservation in western Europe: an Arthropod perspective." Belgian Journal of Entomology **2**: 59-75.

Del Campo, J., M. E. Sieracki, R. Molestina, P. Keeling, R. Massana and I. Ruiz-Trillo (2014). "The others: our biased perspective of eukaryotic genomes." Trends in Ecology & Evolution **29**(5): 252-259.

Ding, Z., Y. Wen, B. Yang, Y. Zhang, S. Liu, Z. Liu and Z. Han (2013). "Biochemical mechanisms of imidacloprid resistance in Nilaparvata lugens: Over-expression of cytochrome P450 CYP6AY1." Insect Biochemistry and Molecular Biology **43**(11): 1021-1027.

Dolédec, S., J. Olivier and B. Statzner (2000). "Accurate description of the abundance of taxa and their biological traits in stream invertebrate communities: effects of taxonomic and spatial resolution." Archiv für Hydrobiologie **148**(1): 25-43.

Donkin, P. (2009). Quantitative Structure-Activity Relationships. Handbook of Ecotoxicology, Blackwell Publishing Ltd.**:** 785-811.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell and S. Lautenbach (2013). "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance." Ecography **36**(1): 27-46.

Dyer, S. D., D. J. Versteeg, S. E. Belanger, J. G. Chaney and F. L. Mayer (2006). "Interspecies Correlation Estimates Predict Protective Environmental Concentrations." Environmental Science & Technology **40**(9): 3102-3111.

Dyer, S. D., D. J. Versteeg, S. E. Belanger, J. G. Chaney, S. Raimondo and M. G. Barron (2008). "Comparison of Species Sensitivity Distributions Derived from Interspecies Correlation Models to Distributions used to Derive Water Quality Criteria." Environmental Science & Technology **42**(8): 3076-3083.

EFSA (2009). "Council directive 91/414/EEC. Azoxystrobin. Report and proposed decision of the United Kingdom made to the European Commission under commission regulation 737/2007. Accessed via http://dar.efsa.europa.eu/dar-web/provision."

EFSA (2014). "Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products." Efsa Journal **12**(3): 3589.

EFSA Panel on Plant Protection Products and their Residues, C. Ockleford, P. Adriaanse, P. Berny, T. Brock, S. Duquesne, S. Grilli, A. F. Hernandez-Jerez, S. H. Bennekou, M. Klein, T. Kuhl, R. Laskowski, K. Machera, O. Pelkonen, S. Pieper, R. H. Smith, M. Stemmer, I. Sundh, A. Tiktak, C. J. Topping, G. Wolterink, N. Cedergreen, S. Charles, A. Focks, M. Reed, M. Arena, A. Ippolito, H. Byers and I. Teodorovic (2018). "Scientific Opinion on the state of the art of Toxicokinetic/Toxicodynamic (TKTD) effect models for regulatory risk assessment of pesticides for aquatic organisms." EFSA Journal **16**(8): e05377.

Eisner, B., J. Doering, S. Beitel and M. Hecker (2019). "Cross-Species Extrapolation Using a Simplified In Vitro Tissue Explant Assay in Fish."

Enoch, S., M. Hewitt, M. Cronin, S. Azam and J. Madden (2008). "Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree." Chemosphere **73**(3): 243-248.

Erickson, R. J. (2013). "The biotic ligand model approach for addressing effects of exposure water chemistry on aquatic toxicity of metals: Genesis and challenges." Environmental Toxicology and Chemistry **32**(6): 1212-1214.

Escher, B. I. and J. L. M. Hermens (2002). "Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects." Environmental Science & Technology **36**(20): 4201-4217.

European Commission (1991). "Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market." Official Journal of the European Communities **230**: 1-32.

European Commission (1998). "Directive 98/8/EC of the European Parliament and of the Council of 16 February 1998 concerning the placing of biocidal products on the market." Official Journal of the European Communities **41**: 123.

European Commission (2000). "Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy." Official journal of the European communities **22**(12): 2000.

European Commission (2007). "Regulation (EC) no. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, evaluation, authorisation and restriction of chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) no. 793/93 and Commission Regulation (EC) no. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC, and 2000/21/EC." The official Journal of the European Union **136**: 3-280.

Farmahin, R., D. Wu, D. Crump, J. C. Hervé, S. P. Jones, M. E. Hahn, S. I. Karchner, J. P. Giesy, S. J. Bursian and M. J. Zwiernik (2012). "Sequence and in vitro function of chicken, ring-necked pheasant, and

Japanese quail AHR1 predict in vivo sensitivity to dioxins." Environmental science & technology **46**(5): 2967-2975.
Federhen, S. (2011). "The NCBI Taxonomy database." Nucleic Acids Research **40**(D1): D136-D143.
Fedorenkova, A., J. A. Vonk, H. R. Lenders, N. J. Ouborg, A. M. Breure and A. J. Hendriks (2010). "Ecotoxicogenomics: Bridging the gap between genes and populations." Environmental science & technology **44**(11): 4328-4333.
Feng, C., F. Wu, S. Dyer, H. Chang and X. Zhao (2013). "Derivation of freshwater quality criteria for zinc using interspecies correlation estimation models to protect aquatic life in China." Chemosphere **90**(3): 1177-1183.
Focks, A., D. Belgers, M.-C. Boerwinkel, L. Buijse, I. Roessink and P. J. Van den Brink (2018). "Calibration and validation of toxicokinetic-toxicodynamic models for three neonicotinoids and some aquatic macroinvertebrates." Ecotoxicology **27**(7): 992-1007.
Franco, A., O. R. Price, S. Marshall, O. Jolliet, P. J. Van den Brink, A. Rico, A. Focks, F. De Laender and R. Ashauer (2016). "Towards refined environmental scenarios for ecological risk assessment of down-the-drain chemicals in freshwater environments." Integrated Environmental Assessment and Management: n/a-n/a.
Friant, S. L. and L. Henry (1985). "Relationship between toxicity of certain organic compounds and their concentrations in tissues of aquatic organisms: A perspective." Chemosphere **14**(11): 1897-1907.
Frimpong, E. A. and P. L. Angermeier (2009). "Fish traits: a database of ecological and life-history traits of freshwater fishes of the United States." Fisheries **34**(10): 487-495.
Froese, R. and D. Pauly (2000). FishBase 2000: concepts designs and data sources, WorldFish.
Gabsi, F. and T. G. Preuss (2014). "Modelling the impact of the environmental scenario on population recovery from chemical stress exposure: A case study using Daphnia magna." Aquatic Toxicology **156**: 221-229.
Gadd, G. M. and A. J. Griffiths (1977). "Microorganisms and heavy metal toxicity." Microbial ecology **4**(4): 303-317.
Gallagher, R. V., D. S. Falster, B. S. Maitner, R. Salguero-Gómez, V. Vandvik, W. D. Pearse, F. D. Schneider, J. Kattge, J. H. Poelen, J. S. Madin, M. J. Ankenbrand, C. Penone, X. Feng, V. M. Adams, J. Alroy, S. C. Andrew, M. A. Balk, L. M. Bland, B. L. Boyle, C. H. Bravo-Avila, I. Brennan, A. J. R. Carthey, R. Catullo, B. R. Cavazos, D. A. Conde, S. L. Chown, B. Fadrique, H. Gibb, A. H. Halbritter, J. Hammock, J. A. Hogan, H. Holewa, M. Hope, C. M. Iversen, M. Jochum, M. Kearney, A. Keller, P. Mabee, P. Manning, L. McCormack, S. T. Michaletz, D. S. Park, T. M. Perez, S. Pineda-Munoz, C. A. Ray, M. Rossetto, H. Sauquet, B. Sparrow, M. J. Spasojevic, R. J. Telford, J. A. Tobias, C. Violle, R. Walls, K. C. B. Weiss, M. Westoby, I. J. Wright and B. J. Enquist (2020). "Open Science principles for accelerating trait-based science across the Tree of Life." Nature Ecology & Evolution **4**(3): 294-303.
Gaston, K. J., K. Charman, S. F. Jackson, P. R. Armsworth, A. Bonn, R. A. Briers, C. S. Callaghan, R. Catchpole, J. Hopkins and W. E. Kunin (2006). "The ecological effectiveness of protected areas: the United Kingdom." Biological Conservation **132**(1): 76-87.
Gayraud, S., B. Statzner, P. Bady, A. Haybachp, F. Schöll, P. Usseglio-Polatera and M. Bacchi (2003). "Invertebrate traits for the biomonitoring of large European rivers: an initial assessment of alternative metrics." Freshwater Biology **48**(11): 2045-2064.
Gergs, A., D. Kulkarni and T. G. Preuss (2015). "Body size-dependent toxicokinetics and toxicodynamics could explain intra-and interspecies variability in sensitivity." Environmental pollution **206**: 449-455.
Gerhardt, A. (1993). "Review of impact of heavy metals on stream invertebrates with special emphasis on acid conditions." Water, Air, and Soil Pollution **66**(3-4): 289-314.
Gerritsen, A., N. van der Hoeven and A. Pielaat (1998). "The Acute Toxicity of Selected Alkylphenols to Young and AdultDaphnia magna." Ecotoxicology and Environmental Safety **39**(3): 227-232.
Gleason, T. R. and D. E. Nacci (2001). "Risks of endocrine-disrupting compounds to wildlife: extrapolating from effects on individuals to population response." Human and Ecological Risk Assessment: An International Journal **7**(5): 1027-1042.
Griffith, D. A. and P. R. Peres-Neto (2006). "Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses." Ecology **87**(10): 2603-2613.
Guénard, G. (2018). A phylogenetic modelling tutorial using Phylogenetic Eigenvector Maps (PEM) as implemented in R package MPSEM (0.3-4).
Guénard, G., P. Legendre and P. Peres-Neto (2013). "Phylogenetic eigenvector maps: a framework to model and predict species traits." Methods in Ecology and Evolution **4**(12): 1120-1131.
Guénard, G., P. C. Ohe, D. de Zwart, P. Legendre and S. Lek (2011). "Using phylogenetic information to predict species tolerances to toxic chemicals." Ecological Applications **21**(8): 3178-3190.
Guénard, G., P. C. von der Ohe, S. C. Walker, S. Lek and P. Legendre (2014). "Using phylogenetic information and chemical properties to predict species tolerances to pesticides." Proceedings of the Royal Society of London B: Biological Sciences **281**(1789): 20133239.
Guha, R. (2007). "Chemical Informatics Functionality in R." Journal of Statistical Software **6**(18).
Hammers-Wirtz, M. and H. T. Ratte (2000). "Offspring fitness in Daphnia: Is the Daphnia reproduction test appropriate for extrapolating effects on the population level?" Environmental Toxicology and Chemistry **19**(7): 1856-1866.
Hartig, F., J. M. Calabrese, B. Reineking, T. Wiegand and A. Huth (2011). "Statistical inference for stochastic simulation models – theory and application." Ecology Letters **14**(8): 816-827.
Hawkins, D. M., S. C. Basak and D. Mills (2003). "Assessing Model Fit by Cross-Validation." Journal of Chemical Information and Computer Sciences **43**(2): 579-586.
He, K. S., B. A. Bradley, A. F. Cord, D. Rocchini, M. N. Tuanmu, S. Schmidtlein, W. Turner, M. Wegmann and N. Pettorelli (2015). "Will remote sensing shape the next generation of species distribution models?" Remote Sensing in Ecology and Conservation **1**(1): 4-18.

Hébert, M.-P., B. E. Beisner and R. Maranger (2016). "A compilation of quantitative functional traits for marine and freshwater crustacean zooplankton." Ecology **97**(4): 1081-1081.
Hebert, P. D., A. Cywinska, S. L. Ball and J. R. Dewaard (2003). "Biological identifications through DNA barcodes." Proceedings of the Royal Society of London. Series B: Biological Sciences **270**(1512): 313-321.
Hendriks, A. J., A. van der Linde, G. Cornelissen and D. T. Sijm (2001). "The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight." Environmental Toxicology and Chemistry: An International Journal **20**(7): 1399-1420.
Hering, D., A. Schmidt-Kloiber, J. Murphy, S. Lücke, C. Zamora-Munoz, M. J. López-Rodríguez, T. Huber and W. Graf (2009). "Potential impact of climate change on aquatic insects: a sensitivity analysis for European caddisflies (Trichoptera) based on distribution patterns and ecological preferences." Aquatic Sciences **71**(1): 3-14.
Hershkovitz, Y., V. Dahm, A. W. Lorenz and D. Hering (2015). "A multi-trait approach for the identification and protection of European freshwater species that are potentially vulnerable to the impacts of climate change." Ecological Indicators **50**: 150-160.
Hickey, C. W. and W. H. Clements (1998). "Effects of heavy metals on benthic macroinvertebrate communities in New Zealand streams." Environmental Toxicology and Chemistry: An International Journal **17**(11): 2338-2346.
Hiley, J. R., R. B. Bradbury, M. Holling and C. D. Thomas (2013). "Protected areas act as establishment centres for species colonizing the UK." Proceedings of the Royal Society B: Biological Sciences **280**(1760): 20122310.
Illies, J. (1978). Limnofauna europaea, Fischer Stuttgart.
Ippolito, A., R. Todeschini and M. Vighi (2012). "Sensitivity assessment of freshwater macroinvertebrates to pesticides using biological traits." Ecotoxicology **21**(2): 336-352.
ITIS. (2019). "Integrated Taxonomic Information System." from http://www.itis.gov.
Jager, T. (2011). Some good reasons to ban EC x and related concepts in ecotoxicology, ACS Publications.
Jager, T. (2012). "Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology." Environmental Toxicology and Chemistry **31**(2): 228-229.
Jager, T. (2019). openGUTS. http://openguts.info/index.html.
Jager, T., C. Albert, T. G. Preuss and R. Ashauer (2011). "General unified threshold model of survival--a toxicokinetic-toxicodynamic framework for ecotoxicology." Environ Sci Technol **45**(7): 2529-2540.
Jager, T., E. H. Heugens and S. A. Kooijman (2006). "Making sense of ecotoxicological test results: towards application of process-based models." Ecotoxicology **15**(3): 305-314.
Jager, T., B. T. Martin and E. I. Zimmer (2013). "DEBkiss or the quest for the simplest generic model of animal life history." J Theor Biol **328**: 9-18.
Johnson, J. B. and K. S. Omland (2004). "Model selection in ecology and evolution." Trends Ecol Evol **19**(2): 101-108.
Karunker, I., J. Benting, B. Lueke, T. Ponge, R. Nauen, E. Roditakis, J. Vontas, K. Gorman, I. Denholm and S. Morin (2008). "Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae)." Insect Biochemistry and Molecular Biology **38**(6): 634-644.
Kase, R., M. Korkaric, I. Werner and M. Ågerstrand (2016). "Criteria for Reporting and Evaluating ecotoxicity Data (CRED): comparison and perception of the Klimisch and CRED methods for evaluating reliability and relevance of ecotoxicity studies." Environmental Sciences Europe **28**(1): 7.
Kienzler, A., M. G. Barron, S. E. Belanger, A. Beasley and M. R. Embry (2017). "Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches." Environ Sci Technol **51**(17): 10203-10211.
Kim, Y. H. and S. H. Lee (2013). "Which acetylcholinesterase functions as the main catalytic enzyme in the Class Insecta?" Insect Biochemistry and Molecular Biology **43**(1): 47-53.
Klaper, R. and M. A. Thomas (2004). "At the Crossroads of Genomics and Ecology: The Promise of a Canary on a Chip." BioScience **54**(5): 403-412.
Koepfli, K.-P., B. Paten, K. C. o. S. Genome and S. J. O'Brien (2015). "The Genome 10K Project: a way forward." Annual review of animal biosciences **3**: 57-111.
Kooijman, S. (1987). "A safety factor for LC50 values allowing for differences in sensitivity among species." Water Research **21**(3): 269-276.
Kretschmann, A., R. Ashauer, T. G. Preuss, P. Spaak, B. I. Escher and J. Hollender (2011). "Toxicokinetic Model Describing Bioconcentration and Biotransformation of Diazinon in Daphnia magna." Environmental Science & Technology **45**(11): 4995-5002.
Ladroue, C., S. Guo, K. Kendrick and J. Feng (2009). "Beyond element-wise interactions: identifying complex interactions in biological processes." PloS one **4**(9).
LaLone, C. A., D. L. Villeneuve, L. D. Burgoon, C. L. Russom, H. W. Helgen, J. P. Berninger, J. E. Tietge, M. N. Severson, J. E. Cavallin and G. T. Ankley (2013). "Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action." Aquatic Toxicology **144-145**: 141-154.
LaLone, C. A., D. L. Villeneuve, D. Lyons, H. W. Helgen, S. L. Robinson, J. A. Swintek, T. W. Saari and G. T. Ankley (2016). "Editor's Highlight: Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity." Toxicol Sci **153**(2): 228-245.
Lamouroux, N., N. L. Poff and P. L. Angermeier (2002). "Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients." Ecology **83**(7): 1792-1807.
Lange, K., C. R. Townsend and C. D. Matthaei (2016). "A trait-based framework for stream algal communities." Ecology and Evolution **6**(1): 23-36.

Larras, F., F. o. Keck, B. Montuelle, F. d. r. Rimet and A. s. Bouchez (2014). "Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring?" Environmental science & technology **48**(3): 1921-1930.

Legendre, P. and L. F. Legendre (2012). Numerical ecology, Elsevier.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, R. Durbin, S. V. Edwards, F. Forest, M. T. P. Gilbert, M. M. Goldstein, I. V. Grigoriev, K. J. Hackett, D. Haussler, E. D. Jarvis, W. E. Johnson, A. Patrinos, S. Richards, J. C. Castilla-Rubio, M.-A. van Sluys, P. S. Soltis, X. Xu, H. Yang and G. Zhang (2018). "Earth BioGenome Project: Sequencing life for the future of life." Proceedings of the National Academy of Sciences **115**(17): 4325-4333.

Liess, M. and M. Beketov (2011). "Traits and stress: keys to identify community effects of low levels of toxicants in test systems." Ecotoxicology **20**(6): 1328-1340.

Liess, M. and P. C. Von der Ohe (2005). "Analyzing effects of pesticides on invertebrate communities in streams." Environmental Toxicology and Chemistry **24**(4): 954-965.

Lumley, T. and A. Miller. (2017). "Leaps: Regression Subset Selection." Retrieved 22-1-2018, 2018, from https://CRAN.R-project.org/package=leaps.

Macaulay, S. J., D. B. Buchwalter and C. D. Matthaei (2020). "Water temperature interacts with the insecticide imidacloprid to alter acute lethal and sublethal toxicity to mayfly larvae." New Zealand Journal of Marine and Freshwater Research **54**(1): 115-130.

Malaj, E., G. Guénard, R. B. Schäfer and P. C. Ohe (2016). "Evolutionary patterns and physicochemical properties explain macroinvertebrate sensitivity to heavy metals." Ecological applications **26**(4): 1249-1259.

Maltby, L., N. Blake, T. C. M. Brock and P. J. Van den Brink (2005). "Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems." Environmental Toxicology and Chemistry **24**(2): 379-388.

Manjon, C., B. J. Troczka, M. Zaworra, K. Beadle, E. Randall, G. Hertlein, K. S. Singh, C. T. Zimmer, R. A. Homem, B. Lueke, R. Reid, L. Kor, M. Kohler, J. Benting, M. S. Williamson, T. G. E. Davies, L. M. Field, C. Bass and R. Nauen (2018). "Unravelling the Molecular Determinants of Bee Sensitivity to Neonicotinoid Insecticides." Current Biology **28**(7): 1137-1143.e1135.

Mannheimer, J. D., D. L. Duval, A. Prasad and D. L. Gustafson (2019). "A systematic analysis of genomics-based modeling approaches for prediction of drug response to cytotoxic chemotherapies." BMC medical genomics **12**(1): 87.

Martin, T., D. Young, C. Lilavois and M. Barron (2015). "Comparison of global and mode of action-based models for aquatic toxicity." SAR and QSAR in Environmental Research **26**(3): 245-262.

Maul, J. D., A. A. Brennan, A. D. Harwood and M. J. Lydy (2008). "Effect of sediment-associated pyrethroids, fipronil, and metabolites on Chironomus tentans growth rate, body mass, condition index, immobilization, and survival." Environmental Toxicology and Chemistry **27**(12): 2582-2590.

Mayer, F. L. and M. R. Ellersieck (1986). Manual of acute toxicity: interpretation and data base for 410 chemicals and 66 species of freshwater animals, US Department of the Interior, Fish and Wildlife Service.

McCarty, L., J. Henry and A. Houston (1978). "Toxicity of cadmium to goldfish, Carassius auratus, in hard and soft water." Journal of the Fisheries Board of Canada **35**(1): 35-42.

McCarty, L., P. Landrum, S. Luoma, J. Meador, A. Merten, B. Shephard and A. van Wezel (2011). "Advancing environmental toxicology through chemical dosimetry: External exposures versus tissue residues." Integrated Environmental Assessment and Management **7**(1): 7-27.

McCarty, L. and K. Munkittrick (1996). "Environmental biomarkers in aquatic toxicology: fiction, fantasy, or functional?".

McClellan-Green, P., J. Romano and E. Oberdörster (2007). "Does gender really matter in contaminant exposure? A case study using invertebrate models." Environmental Research **104**(1): 183-191.

Médail, F. and P. Quézel (1999). "Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities." Conservation Biology **13**(6): 1510-1513.

Mendiburu, F. d. (2017). agricolae: Statistical Procedures for Agricultural Research. **version 1.2-8**.

Mittermeier, R. A., N. Myers, J. B. Thomsen, G. A. Da Fonseca and S. Olivieri (1998). "Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities." Conservation biology **12**(3): 516-520.

Moermond, C. T., R. Kase, M. Korkaric and M. Ågerstrand (2016). "CRED: Criteria for reporting and evaluating ecotoxicity data." Environmental Toxicology and Chemistry **35**(5): 1297-1309.

Moog, O., A. Schmidt-Kloiber, T. Ofenböck and J. Gerritsen (2004). Does the ecoregion approach support the typological demands of the EU 'Water Framework Directive'? Integrated assessment of running waters in Europe, Springer**:** 21-33.

Moreno-Hagelsieb, G. and K. Latimer (2007). "Choosing BLAST options for better detection of orthologs as reciprocal best hits." Bioinformatics **24**(3): 319-324.

Morrison, B. (1979). "An investigation into the effects of the piscicide antimycin A on the fish and invertebrates of a Scottish stream." Aquaculture Research **10**(3): 111-122.

Morrissey, C. A., P. Mineau, J. H. Devries, F. Sanchez-Bayo, M. Liess, M. C. Cavallaro and K. Liber (2015). "Neonicotinoid contamination of global surface waters and associated risk to aquatic invertebrates: a review." Environment international **74**: 291-303.

Murphy, P. (1978). "The temporal variability in biotic indices." Environmental Pollution (1970) **17**(3): 227-236.

Nagai, T. (2016). "Ecological effect assessment by species sensitivity distribution for 68 pesticides used in Japanese paddy fields." Journal of Pesticide Science **41**(1): 6-14.

Nelson, D. L. and M. M. Cox (2008). Lehninger principles of biochemistry. New York - Basingstoke, W.H. Freeman and Company.

Nendza, M. and M. T. Muller (2000). "Discriminating toxicant classes by MOA: 2. Physico-Chemical Descriptors." Quantitative Structure-Activity Relationships **19**: 581-598.

Nyman, A.-M., K. Schirmer and R. Ashauer (2012). "Toxicokinetic-toxicodynamic modelling of survival of Gammarus pulex in multiple pulse exposures to propiconazole: model assumptions, calibration data requirements and predictive power." Ecotoxicology **21**(7): 1828-1840.

OECD (2006). Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q) SAR] Models in the Assessment of New and Existing Chemicals, Organisation of Economic Cooperation and Development. OECD Series on Testing ….

OECD (2019). Test No. 203: Fish, Acute Toxicity Test.

OECD, E. D. (2004). Report from the Expert group on quantitative structure–activity relationships (Q) SAR on the principles of validation of (Q) SARs, Organisation for Economic Cooperation and Development Paris,, France**:** 206.

Pacyna, J. M., K. Breivik, J. Münch and J. Fudala (2003). "European atmospheric emissions of selected persistent organic pollutants, 1970–1995." Atmospheric Environment **37**: 119-131.

Pacyna, J. M., E. G. Pacyna, F. Steenhuisen and S. Wilson (2003). "Mapping 1995 global anthropogenic emissions of mercury." Atmospheric Environment **37**: 109-117.

Pagel, M. (1999). "Inferring the historical patterns of biological evolution." Nature **401**: 877.

Paquin, P. R., J. W. Gorsuch, S. Apte, G. E. Batley, K. C. Bowles, P. G. C. Campbell, C. G. Delos, D. M. Di Toro, R. L. Dwyer, F. Galvez, R. W. Gensemer, G. G. Goss, C. Hogstrand, C. R. Janssen, J. C. McGeer, R. B. Naddy, R. C. Playle, R. C. Santore, U. Schneider, W. A. Stubblefield, C. M. Wood and K. B. Wu (2002). "The biotic ligand model: a historical overview." Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **133**(1): 3-35.

Pascoe, D., S. A. Evans and J. Woodworth (1986). "Heavy metal toxicity to fish and the influence of water hardness." Archives of Environmental Contamination and Toxicology **15**(5): 481-487.

Pebesma, E. J. and R. S. Bivand (2005). "Classes and methods for spatial data in R." R news **5**(2): 9-13.

Pennie, B., H.-J. Ahr, C. Braun, R. Bars, E. Jacob and F. Carpanini (2001). Genomics, Transcript Profiling,Proteomics and Metabonomics (GTPM) An Introduction (white paper). Brussels, ECETOC.

Pilière, A. F. H., W. C. E. P. Verberk, M. Gräwe, A. M. Breure, S. D. Dyer, L. Posthuma, D. de Zwart, M. A. J. Huijbregts and A. M. Schipper (2016). "On the importance of trait interrelationships for understanding environmental responses of stream macroinvertebrates." Freshwater Biology **61**(2): 181-194.

Poff, N. L., J. D. Olden, N. K. Vieira, D. S. Finn, M. P. Simmons and B. C. Kondratieff (2006). "Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships." Journal of the North American Benthological Society **25**(4): 730-755.

Posthuma, L., G. W. Suter II and T. P. Traas (2001). Species sensitivity distributions in ecotoxicology, CRC press.

Poteat, M. D. and D. B. Buchwalter (2014). "Phylogeny and size differentially influence dissolved Cd and Zn bioaccumulation parameters among closely related aquatic insects." Environmental science & technology **48**(9): 5274-5281.

Poteat, M. D., L. M. Jacobus and D. B. Buchwalter (2015). "The importance of retaining a phylogenetic perspective in traits-based community analyses." Freshwater biology **60**(7): 1330-1339.

Poynton, H. C., W. E. Robinson, B. J. Blalock and R. E. Hannigan (2014). "Correlation of transcriptomic responses and metal bioaccumulation in Mytilus edulis L. reveals early indicators of stress." Aquatic Toxicology **155**: 129-141.

Prevedouros, K., M. MacLeod, K. C. Jones and A. J. Sweetman (2004). "Modelling the fate of persistent organic pollutants in Europe: parameterisation of a gridded distribution model." Environmental Pollution **128**(1-2): 251-261.

R Core Team (2018). "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012." URL http://www. R-project. org.

R Core Team (2020). "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012." URL http://www. R-project. org.

Raby, M., M. Nowierski, D. Perlov, X. Zhao, C. Hao, D. G. Poirier and P. K. Sibley (2018). "Acute toxicity of 6 neonicotinoid insecticides to freshwater invertebrates." Environmental Toxicology and Chemistry **37**(5): 1430-1445.

Rahbek, C., M. K. Borregaard, A. Antonelli, R. K. Colwell, B. G. Holt, D. Nogues-Bravo, C. M. Rasmussen, K. Richardson, M. T. Rosing and R. J. Whittaker (2019). "Building mountain biodiversity: Geological and evolutionary processes." Science **365**(6458): 1114-1119.

Rahbek, C., M. K. Borregaard, R. K. Colwell, B. Dalsgaard, B. G. Holt, N. Morueta-Holme, D. Nogues-Bravo, R. J. Whittaker and J. Fjeldså (2019). "Humboldt's enigma: What causes global patterns of mountain biodiversity?" Science **365**(6458): 1108-1113.

Raimondo, S. and M. Barron (2019). "Application of Interspecies Correlation Estimation (ICE) models and QSAR in estimating species sensitivity to pesticides." SAR and QSAR in Environmental Research: 1-18.

Raimondo, S., C. R. Jackson and M. G. Barron (2010). "Influence of taxonomic relatedness and chemical mode of action in acute interspecies estimation models for aquatic species." Environmental science & technology **44**(19): 7711-7716.

Raimondo, S., C. R. Lilavois and M. G. Barron (2015). "Web-based interspecies correlation estimation (Web-ICE) for acute toxicity: user manual. Version 3.3." Office of Research and Development. US Environmental Protection Agency, Gulf Breeze, FL, USA.

Raimondo, S., P. Mineau and M. G. Barron (2007). "Estimation of Chemical Toxicity to Wildlife Species Using Interspecies Correlation Models." Environmental Science & Technology **41**(16): 5888-5894.

Raimondo, S., D. N. Vivian and M. G. Barron (2010). "Web-based interspecies correlation estimation (Web-ICE) for acute toxicity: user manual. Version 3.0." Office of Research and Development. US Environmental Protection Agency, Gulf Breeze, FL, USA.

Räsänen, K., A. Laurila and J. Merilä (2003). "Geographic variation in acid stress tolerance of the Moor frog, Rana arvalis. I. Local Adaptation." Evolution **57**(2): 352-362.

Rasmussen, J. J., P. Wiberg-Larsen, E. A. Kristensen, N. Cedergreen and N. Friberg (2013). "Pyrethroid effects on freshwater invertebrates: A meta-analysis of pulse exposures." Environmental Pollution **182**: 479-485.

Reinert, K. H., J. M. Giddings and L. Judd (2002). "Effects analysis of time-varying or repeated exposures in aquatic ecological risk assessment of agrochemicals." Environmental Toxicology and Chemistry **21**(9): 1977-1992.

Reynolds, C. S., V. Huszar, C. Kruk, L. Naselli-Flores and S. Melo (2002). "Towards a functional classification of the freshwater phytoplankton." Journal of plankton research **24**(5): 417-428.

Rico, A. and P. J. Van den Brink (2015). "Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action." Environ Toxicol Chem **34**(8): 1907-1917.

Rico, A., P. J. Van den Brink, R. Gylstra, A. Focks and T. C. M. Brock (2016). "Developing ecological scenarios for the prospective aquatic risk assessment of pesticides." Integrated Environmental Assessment and Management **12**(3): 510-521.

Rider, C. V., G. E. Dinse, D. M. Umbach, J. E. Simmons and R. C. Hertzberg (2018). Predicting Mixture Toxicity with Models of Additivity. Chemical Mixtures and Combined Chemical and Nonchemical Stressors: Exposure, Toxicity, Analysis, and Risk. C. V. Rider and J. E. Simmons. Cham, Springer International Publishing**:** 235-270.

Rinkevich, F. D., Y. Du and K. Dong (2013). "Diversity and convergence of sodium channel mutations involved in resistance to pyrethroids." Pesticide biochemistry and physiology **106**(3): 93-100.

Rinkevich, F. D., C. Su, T. A. Lazo, D. J. Hawthorne, W. M. Tingey, S. Naimov and J. G. Scott (2012). "Multiple evolutionary origins of knockdown resistance (kdr) in pyrethroid-resistant Colorado potato beetle, Leptinotarsa decemlineata." Pesticide biochemistry and physiology **104**(3): 192-200.

Rivera-Ingraham, G. A. and J.-H. Lignot (2017). "Osmoregulation, bioenergetics and oxidative stress in coastal marine invertebrates: raising the questions for future research." Journal of Experimental Biology **220**(10): 1749-1760.

Roessink, I., L. B. Merga, H. J. Zweers and P. J. Van den Brink (2013). "The neonicotinoid imidacloprid shows high chronic toxicity to mayfly nymphs." Environmental Toxicology and Chemistry **32**(5): 1096-1100.

Rohr, J. R., C. J. Salice and R. M. Nisbet (2016). "The pros and cons of ecological risk assessment based on data from different levels of biological organization." Critical Reviews in Toxicology **46**(9): 756-784.

Rosenfeld, J. S. (2002). "Functional redundancy in ecology and conservation." Oikos **98**(1): 156-162.

Rotvit, L. and D. Jacobsen (2013). "Temperature increase and respiratory performance of macroinvertebrates with different tolerances to organic pollution." Limnologica **43**(6): 510-515.

Rubach, M. N., R. Ashauer, D. B. Buchwalter, H. J. De Lange, M. Hamer, T. G. Preuss, K. Töpke and S. J. Maund (2011). "Framework for traits-based assessment in ecotoxicology." Integrated environmental assessment and management **7**(2): 172-186.

Rubach, M. N., R. Ashauer, S. J. Maund, D. J. Baird and P. J. Van den Brink (2010). "Toxicokinetic variation in 15 freshwater arthropod species exposed to the insecticide chlorpyrifos." Environmental Toxicology and Chemistry **29**(10): 2225-2234.

Rubach, M. N., D. J. Baird, M.-C. Boerwinkel, S. J. Maund, I. Roessink and P. J. Van den Brink (2012). "Species traits as predictors for intrinsic sensitivity of aquatic invertebrates to the insecticide chlorpyrifos." Ecotoxicology **21**(7): 2088-2101.

Rubach, M. N., D. J. Baird and P. J. Van den Brink (2010). "A new method for ranking mode-specific sensitivity of freshwater arthropods to insecticides and its relationship to biological traits." Environmental Toxicology and Chemistry **29**(2): 476-487.

Rubach, M. N., S. J. H. Crum and P. J. Van den Brink (2011). "Variability in the Dynamics of Mortality and Immobility Responses of Freshwater Arthropods Exposed to Chlorpyrifos." Archives of Environmental Contamination and Toxicology **60**(4): 708-721.

Sayers, E. W., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, E. Yaschenko and J. Ye (2009). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res **37**(Database issue): D5-15.

Schäfer, R. B., B. J. Kefford, L. Metzeling, M. Liess, S. Burgert, R. Marchant, V. Pettigrove, P. Goonan and D. Nugegoda (2011). "A trait database of stream invertebrates for the ecological risk assessment of single and combined effects of salinity and pesticides in South-East Australia." Science of the Total Environment **409**(11): 2055-2063.

Schmidt-Kloiber, A. and D. Hering (2015). "www.freshwaterecology.info – An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences." Ecological Indicators **53**: 271-282.

Schmitt, T. and Z. Varga (2012). "Extra-Mediterranean refugia: the rule and not the exception?" Frontiers in Zoology **9**(1): 22.

Schroer, A. F. W., J. D. M. Belgers, C. M. Brock, A. M. Matser, S. J. Maund and P. J. Van den Brink (2004). "Comparison of Laboratory Single Species and Field Population-Level Effects ofthe Pyrethroid Insecticide λ-Cyhalothrin on Freshwater Invertebrates." Archives of Environmental Contamination and Toxicology **46**(3): 324-335.

Schultz, T. W. and M. T. Cronin (2003). "Essential and desirable characteristics of ecotoxicity quantitative structure–activity relationships." Environmental Toxicology and Chemistry: An International Journal **22**(3): 599-607.

Segner, H. (2011). Chapter 86 - Reproductive and developmental toxicity in fishes. Reproductive and Developmental Toxicology. R. C. Gupta. San Diego, Academic Press**:** 1145-1166.

Segner, H. (2011). "Moving beyond a descriptive aquatic toxicology: the value of biological process and trait information." Aquat Toxicol **105**(3-4 Suppl): 50-55.
Simões, T., S. C. Novais, T. Natal-da-Luz, M. Renaud, S. Leston, F. Ramos, J. Römbke, D. Roelofs, N. M. van Straalen, J. P. Sousa and M. F. L. Lemos (2019). "From laboratory to the field: Validating molecular markers of effect in Folsomia candida exposed to a fungicide-based formulation." Environment International **127**: 522-530.
Snape, J. R., S. J. Maund, D. B. Pickford and T. H. Hutchinson (2004). "Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology." Aquatic toxicology **67**(2): 143-154.
Society of Environmental Toxicology and Chemistry (2019). "Technical Issue Paper: Recommended Minimum Reporting Information for Environmental Toxicity Studies. ."
Spellerberg, I. F. and P. J. Fedor (2003). "A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index." Global Ecology and Biogeography **12**(3): 177-179.
Stemmler, I. and G. Lammel (2009). "Cycling of DDT in the global environment 1950–2002: World ocean returns the pollutant." Geophysical Research Letters **36**(24).
Suter, G. W. (2016). Ecological risk assessment, CRC press.
Tachet, H., P. Richoux, M. Bournaud and P. Usseglio-Polatera (2000). Invertébrés d'eau douce: systématique, biologie, écologie, CNRS éditions Paris.
Ter Braak, C. J. F. (1995). Ordination. Data analysis in community and landscape ecology. Cambridge, Cambridge University Press**:** 9-173.
Thomas, C. D., P. K. Gillingham, R. B. Bradbury, D. B. Roy, B. J. Anderson, J. M. Baxter, N. A. Bourn, H. Q. Crick, R. A. Findon and R. Fox (2012). "Protected areas facilitate species' range expansions." Proceedings of the National Academy of Sciences **109**(35): 14063-14068.
United Nations (2019). World Population Prospects 2019: Data Booklet (ST/ESA/SER. A/424). P. D. Department of Economic Social Affairs, United Nations New York, NY.
USEPA (2002). A review of the reference dose and reference concentration processes, EPA/630/P-02.
USEPA. (2017, 03-14-2017). "ECOTOX User Guide: ECOTOXicology Knowledgebase System. ." Version 4. Retrieved 03-30-2017, 2017.
USEPA (2018). Estimation Programs Interface Suite for Microsoft Windows. U. S. E. P. Agency. Washington, DC, USA.
USEPA (2019). "ECOTOX User Guide: ECOTOXicology Knowledgebase System. Version 5.0. Available: http:/www.epa.gov/ecotox/ (accessed 29-11-2019)."
Usseglio-Polatera, P., M. Bournaud, P. Richoux and H. Tachet (2000). "Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits." Freshwater Biology **43**(2): 175-205.
Van den Berg, S. (2019). Supporting Information for the manuscript entitled "Potential impact of chemical stress on freshwater invertebrates: A sensitivity analysis on continental and national scale based on distribution patterns, biological traits, and lineage". Figshare.
Van den Berg, S. J. P., H. Baveco, E. Butler, F. De Laender, A. Focks, A. Franco, C. Rendal and P. J. Van den Brink (2019). "Modeling the Sensitivity of Aquatic Macroinvertebrates to Chemicals Using Traits." Environmental Science & Technology **53**(10): 6025-6034.
Van den Brink, P. J. (2008). "Ecological Risk Assessment: From Book-Keeping to Chemical Stress Ecology." Environmental Science & Technology **42**(24): 8999-9004.
Van den Brink, P. J., A. C. Alexander, M. Desrosiers, W. Goedkoop, P. L. Goethals, M. Liess and S. D. Dyer (2011). "Traits-based approaches in bioassessment and ecological risk assessment: Strengths, weaknesses, opportunities and threats." Integrated environmental assessment and management **7**(2): 198-208.
Van den Brink, P. J., D. M. Buijert - de Gelder, T. C. M. Brock, I. Roessink and A. Focks (2019). "Exposure pattern-specific species sensitivity distributions for the ecological risk assessments of insecticides." Ecotoxicology and Environmental Safety **180**: 252-258.
Van den Brink, P. J., J. M. Van Smeden, R. S. Bekele, W. Dierick, D. M. De Gelder, M. Noteboom and I. Roessink (2016). "Acute and chronic toxicity of neonicotinoids to nymphs of a mayfly species and some notes on seasonal differences." Environmental Toxicology and Chemistry **35**(1): 128-133.
Van der Lee, G. H., M. H. S. Kraak, R. C. M. Verdonschot and P. F. M. Verdonschot (2020). "Persist or perish: critical life stages determine the sensitivity of invertebrates to disturbances." Aquatic Sciences **82**(2): 24.
van Wezel, A. P., D. A. M. de Vries, S. Kostense, D. T. H. M. Sijm and A. Opperhuizen (1995). "Intraspecies variation in lethal body burdens of narcotic compounds." Aquatic Toxicology **33**(3): 325-342.
Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell and C. E. Cushing (1980). "The river continuum concept." Canadian journal of fisheries and aquatic sciences **37**(1): 130-137.
Verberk, W., C. Van Noordwijk and A. Hildrew (2013). "Delivering on a promise: integrating species traits to transform descriptive community ecology into a predictive science."
Villeneuve, D. L. and N. Garcia-Reyero (2011). "Vision & strategy: Predictive ecotoxicology in the 21st century." Environmental toxicology and chemistry **30**(1): 1-8.
Weston, D. P., H. C. Poynton, G. A. Wellborn, M. J. Lydy, B. J. Blalock, M. S. Sepulveda and J. K. Colbourne (2013). "Multiple origins of pyrethroid insecticide resistance across the species complex of a nontarget aquatic crustacean, Hyalella azteca." Proceedings of the National Academy of Sciences **110**(41): 16532-16537.
Wezel, A. P. v. and A. Opperhuizen (1995). "Narcosis due to environmental pollutants in aquatic organisms: residue-based toxicity, mechanisms, and membrane burdens." Critical Reviews in Toxicology **25**(3): 255-279.

Wiberg-Larsen, P., D. Graeber, E. A. Kristensen, A. Baattrup-Pedersen, N. Friberg and J. J. Rasmussen (2016). "Trait characteristics determine pyrethroid sensitivity in nonstandard test species of freshwater macroinvertebrates: a reality check." Environmental science & technology **50**(10): 4971-4978.
Wijngaarden, R. P. V., T. C. Brock and P. J. Van Den Brink (2005). "Threshold levels for effects of insecticides in freshwater ecosystems: a review." Ecotoxicology **14**(3): 355.
Willming, M. M., C. R. Lilavois, M. G. Barron and S. Raimondo (2016). "Acute toxicity prediction to threatened and endangered species using Interspecies Correlation Estimation (ICE) models." Environmental science & technology **50**(19): 10700-10707.
Windsor, F. M., S. J. Ormerod and C. R. Tyler (2018). "Endocrine disruption in aquatic systems: up-scaling research to address ecological consequences." Biological Reviews **93**(1): 626-641.
Wintle, B. A., M. A. McCarthy, C. T. Volinsky and R. P. Kavanagh (2003). "The use of Bayesian model averaging to better represent uncertainty in ecological models." Conservation Biology **17**(6): 1579-1590.
Worth, A. and G. Patlewicz (2007). "A Compendium of Case Studies that helped to shape the REACH guidance on chemical categories and read across." European Commission report EUR **22481**.
Worth, A. P. (2010). The role of QSAR methodology in the regulatory assessment of chemicals. Recent Advances in QSAR Studies, Springer**:** 367-382.
Wright, J. F. (1994). "Development of RIVPACS in the UK and the value of the underlying data-base." Limnetica **10**(1): 15-31.
Yang, J., X. Zhang, Y. Xie, C. Song, J. Sun, Y. Zhang, J. P. Giesy and H. Yu (2017). "Ecogenomics of Zooplankton Community Reveals Ecological Threshold of Ammonia Nitrogen." Environmental Science & Technology **51**(5): 3057-3064.
Zafar, M. I., J. D. M. Belgers, R. P. A. Van Wijngaarden, A. Matser and P. J. Van den Brink (2012). "Ecological impacts of time-variable exposure regimes to the fungicide azoxystrobin on freshwater communities in outdoor microcosms." Ecotoxicology **21**(4): 1024-1038.
Zhang, X., P. Xia, P. Wang, J. Yang and D. J. Baird (2018). Omics advances in ecotoxicology, ACS Publications.

Summary

The aim of this thesis was to improve the development of models describing species sensitivity towards chemical stressors. This included unravelling decisions that might be of importance in the modelling process, obtaining a better mechanistic understanding of differences in species sensitivity, and providing recommendations for applying such modelling approaches within or across different taxonomic groups of freshwater ecosystems all over the globe. To accomplish this global applicability, we primarily focus on traits-based instead of taxonomy-based approaches on the development of predictive models. A trait is a phenotypic or ecological character of an organism at individual or population level, and describes the physical characteristics, ecological niche and functional role of a species within the ecosystem. We make use of traits, because they are transferable across geographies, add mechanistic and diagnostic knowledge and can be easily translated from taxonomic analyses priory preformed.

We start this thesis with presenting a new predictive modelling approach for potential use in environmental risk assessment (ERA). This new approach constructs macroinvertebrate sensitivity rankings, and subsequently, predictive trait models for a set of pre-defined modes of action (MOA). Each model reveals different taxonomic patterns of species sensitivity, demonstrating that there is not one (set of) species which is most sensitive to all chemicals. The traits included in the models additionally provide information on the mechanisms underlying differences in species sensitivity. We identified, for instance, traits like life cycle duration and respiration mode as important in explaining differences in species sensitivity, and which are hypothesized to be indicators of respectively metabolism and uptake rates. Additionally, we provide an accurate picture of which species-chemical combinations still lack sufficient data, and found that no or incomplete traits data were available for 71% of the species, making the lack of trait data the main obstacle in model construction.

Although these models delivered a new understanding of species sensitivity, they also demonstrate that using traits-data alone will never results in explaining all differences in species sensitivity. Therefore, we develop the models further by adding taxonomic relatedness as an additional predictor of species sensitivity. We indeed find that model performance increases, although only marginally. Besides improving the models further, we also develop the tool further, so that it is now able to predict the sensitivity of species never tested before. We apply the developed models and tool to community assemblages at two different spatial scales (continental and national) and for two different MOAs (narcosis and acetylcholinesterase (AChE) inhibition). The community composition of European freshwater ecoregions is used for the application of our models at the continental scale, while the reference database of the RIVPACS (River InVertebrate Prediction And Classification System) tool is used for river-type scale within the United Kingdom. We found that on a relative scale, 46% and 33% of European species were ranked as more sensitive towards narcosis and AChE inhibition, respectively.

These more sensitive species were distributed with higher occurrences in the south and north-eastern regions, reflecting known continental patterns of endemic macroinvertebrate biodiversity. At the UK level, we found contradicting sensitivity patterns depending on the MOA, with more species displaying relative sensitivity to narcotic MOA in north and north-western regions, and more species with relative sensitivity to AChE inhibition MOA in south and south-western regions.

Next, instead of looking for alternative predictors of species sensitivity, we look for alternative descriptors of species sensitivity. Recently, mechanistic effect models have been suggested as an alternative to the statistical summaries (e.g. LC50s) that are currently being used to assess the potential risk of chemicals to the diversity of all living species. Examples of such quantitative mechanistic effect models are the toxicokinetic- toxicodynamic models of the General Unified Threshold models of Survival (GUTS) framework, which link external exposure and survival effects by describing the processes of uptake, biotransformation, elimination, damage and internal recovery. Parameterization of these models by means of traits enables prediction of standard sensitivity endpoints for a wide range of species and for multiple exposure patterns. We find that GUTS models parameterized on traits are indeed able to approximate the sensitivity of freshwater arthropods, and additionally allow us to understand the underlying processes of species sensitivity better.

Since so many decisions must be made regarding the predictors or descriptors that can be used to describe species sensitivity, we provide a review on currently existing cross-species extrapolation methodologies. We argue that risk assessment can benefit most from modelling approaches when species sensitivity is described based on effects that are ecologically relevant and robust. Additionally, specific attention should be paid to the heterogeneity of the training data, since this strongly influences the reliability of the resulting models. Regarding which predictors are useful for explaining differences in species sensitivity, we develop an integrated framework, combining the strengths of interspecies-correlation, relatedness-based, traits-based, and genomic-based extrapolation methods. Finally, we discuss that regardless of the statistical method used, statistically significant models can be found. However, the usefulness and applicability of these models varies considerably according to modelling choices made. We therefore recommend the publication of scientific code along with scientific studies to simultaneously clarify modelling choices, and enable continuation and elaboration on existing work.

We conclude this thesis by explaining the three ways in which cross-species extrapolation methods can be used in the prospective risk assessment for chemicals: i) to support priority setting procedures by narrowing down further assessment work, ii) to supplement the use of experimental data in weight-of-evidence approaches, and iii) to replace or completely substitute the need for experimental data. Through this development, our approach can help reduce animal testing and contribute towards a new predictive ecotoxicology framework.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Sanne Johanna Petronella van den Berg

born on 14 April 1991 in ‘s Hertogenbosch, the Netherlands

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Wageningen, 16 June 2020

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. dr. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Sanne Johanna Petronella van den Berg** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 41.8 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2016)
- Research in context activity, winner of a Unilever Science Communication Award: 'Create video explaining Modelling decisions in macroinvertebrate chemical sensitivity modelling' (2018)

Other PhD and Advanced MSc Courses

- Dynamic modelling of toxic effects, University of Copenhagen (2016)
- Introduction to agent based models in ecology using Netlogo, Transmitting Science (2016)
- Spatial ecotoxicology and ecotoxicological risk assessment using an open community approach, University of Koblenz-Landau (2017)
- The use of trait based approaches in community ecology and stress ecology, University of Coimbra (2017)
- Presenting with impact, Wageningen Graduate Schools (2017)
- Scientific writing, Wageningen Graduate Schools (2017)
- Data management planning, Wageningen University (2017)
- High Performance Cluster course, Wageningen Universities (2017)
- Introduction to Git and Python, Wageningen University (2018)

Management and Didactic Skills Training

- Supervising three MSc students with thesis and one BSc student (2016-2019)
- Teaching in the MSc courses 'CSERA' (2016=2017) and the BSc/MSc course 'Aquatic Ecology' (2018)

Selection of Oral Presentations

- Ecological scenarios for the prospective risk assessment of aquatic ecosystems. SETAC, 7-11 May 2017, Brussels, Belgium
- Ecological scenarios for the prospective risk assessment of aquatic ecosystems. Symposium for European Freshwater Sciences, 2-7 July 2017, Olomouc, Czech Republic
- Modelling decisions in macroinvertebrate chemical sensitivity modelling, SETAC, 13-17 May, 2018, Rome, Italy
- The combined potential of species-traits and mechanistic-effect models, SETAC, 26-30 May 2019, Helsinki, Finland

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

The research described in this thesis was carried out under the Development of ecological archetypes and models for use in chemical risk assessment, financially supported by Unilever. Printing of this thesis was financially supported by the Aquatic Ecology and Water Quality Management Group of Wageningen University.

Cover design and artwork: Bregje Jaspers || Proefschriftontwerp

Layout: Sanne van den Berg & Proefschriftmaken

Printed by: Proefschriftmaken