



Review

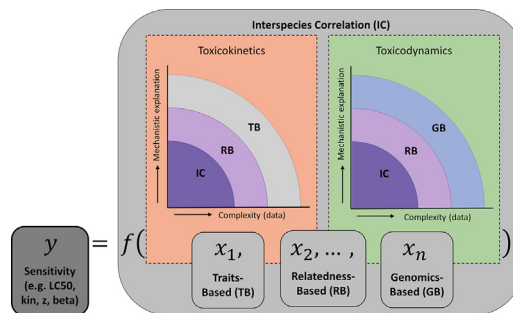
Cross-species extrapolation of chemical sensitivity

Sanne J.P. van den Berg^{a,b,*}, Lorraine Maltby^c, Tom Sinclair^c, Ruoyu Liang^c, Paul J. van den Brink^{a,d}^a Aquatic Ecology and Water Quality Management group, Wageningen University and Research, P.O. box 47, 6700 AA Wageningen, the Netherlands^b Research Unit of Environmental and Evolutionary Biology, Namur Institute of Complex Systems, Institute of Life, Earth, and the Environment, University of Namur, Rue de Bruxelles 61, 5000 Namur, Belgium^c Department of Animal and Plant Sciences, The University of Sheffield, Alfred Denny Building, Western Bank, S10 2TN Sheffield, United Kingdom^d Wageningen Environmental Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands

HIGHLIGHTS

- Methods for the cross-species extrapolation of chemical sensitivity were overviewed.
- Various descriptors of species sensitivity were surveyed.
- Relatedness-, traits-, and genomic-predictors added mechanistic information.
- An integrated framework combining approaches is suggested.
- Statistical considerations important when extrapolating sensitivity are described.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 19 June 2020

Received in revised form 15 August 2020

Accepted 17 August 2020

Available online 19 August 2020

Editor: Henner Hollert

Keywords:

Chemical sensitivity

Cross-species extrapolation

Traits

Genomics

Relatedness

Interspecies correlation

ABSTRACT

Ecosystems are usually populated by many species. Each of these species carries the potential to show a different sensitivity towards all of the numerous chemical compounds that can be present in their environment. Since experimentally testing all possible species-chemical combinations is impossible, the ecological risk assessment of chemicals largely depends on cross-species extrapolation approaches. This review overviews 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 sensitivity is described based on ecologically relevant and robust effects. Additionally, specific attention should be paid to heterogeneity of the training data (e.g. exposure duration, pH, temperature), since this strongly influences the reliability of the resulting models. Regarding which predictors are useful for explaining differences in species sensitivity, we review interspecies-correlation, relatedness-based, traits-based, and genomic-based extrapolation methods, describing 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 develop a conceptual 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 varies considerably. We therefore recommend publication of scientific code along with scientific studies to simultaneously clarify modelling choices and enable elaboration on existing work. In general, this review specifies the data requirements of different cross-species extrapolation methods, aiming to make regulators and publishers more aware that access

* Corresponding author at: Aquatic Ecology and Water Quality Management Group, Wageningen University and Research, P.O. box 47, 6700 AA Wageningen, the Netherlands.
E-mail address: sannejpvandenberg@gmail.com (S.J.P. van den Berg).

to raw- and meta-data needs to be improved to make future cross-species extrapolation efforts successful, enabling their integration into the regulatory environment.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

| | |
|--|----|
| 1. Introduction | 2 |
| 2. How can we describe species sensitivity? | 3 |
| 2.1. Effects | 3 |
| 2.2. Exposure duration | 3 |
| 2.3. Additional selection criteria | 4 |
| 2.4. Units | 5 |
| 2.5. Grouping data, and its effects on explained variance | 5 |
| 3. Which independent variables are useful for explaining differences in species sensitivity? | 5 |
| 3.1. Overview of methods | 5 |
| 3.2. Comparison of methods | 6 |
| 3.2.1. Mechanistic explanation | 6 |
| 3.2.2. Data demand | 7 |
| 3.2.3. Protection of ecological entities | 7 |
| 3.3. A combined approach to predicting sensitivity. | 8 |
| 4. Which statistical considerations are important when extrapolating species sensitivity? | 9 |
| 5. Concluding remarks | 10 |
| Declaration of competing interest | 11 |
| Acknowledgements | 11 |
| References. | 11 |

1. Introduction

An ecosystem generally consists of a diverse species assemblage. Each of the species present in such an 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. Biggs et al., 2007; Clements and Rohr, 2009; Hickey and Clements, 1998). 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 often defined by combining results of single species toxicity tests with assessment factors (e.g. Brock 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) (Brown et al., 2017). 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 extrapolate existing toxicity data to untested organisms. By predicting sensitivity values for a wide range of species, these methods can account for the part of the spatial-temporal variation in species sensitivity that is due to differences in species assemblages within and between sites (e.g. Malaj et al., 2016; Raimondo and Barron, 2019; Van den Berg et al., 2019). However, although several predictive methods have been developed over 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 (Janardan et al., 1984; Mayer and Ellersieck, 1986). A software program to predict acute effects on aquatic and terrestrial species using ICE was developed in the 2000s (Asfaw et al., 2003) and a web-based model is available as Web-ICE (Raimondo et al., 2015). The method has gained popularity for the derivation of water quality criteria (e.g. Dyer et al., 2008; Feng et al., 2013), for example within the WFD (Water Framework Directive, European Commission, 2000).

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 PPR Panel (Panel on Plant Protection Products and their Residues) 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.

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 , Fig. 1). Concerning the cross-species extrapolation

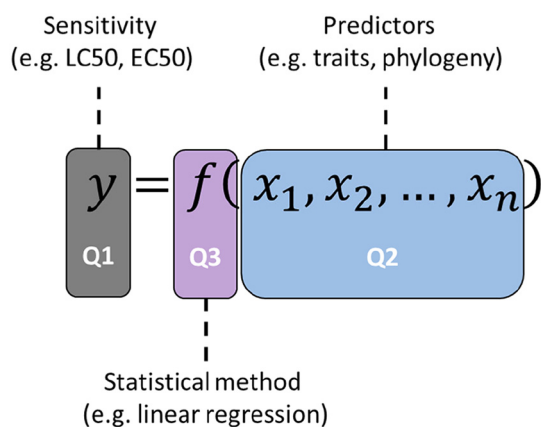


Fig. 1. Schematic overview of the elements making up predictive models. The Qs indicate the elements covered by sub-question 1, 2, and 3 posed in this review.

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, 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 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 identify the range of 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 a conceptual framework. Since our background and expertise lies primarily in the field of aquatic ecotoxicology, most examples mentioned in this review will refer to the aquatic ecosystem. However, the general concepts and theories described and discussed can be applied to any cross-species extrapolation effort.

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 will only be capable of predicting effects on mortality. We will discuss important selection criteria in Sections 2.1–2.4. 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). This will be discussed in more detail in Section 2.5.

2.1. Effects

Effects on mortality are most frequently incorporated into predictive models (Table 1). This is primarily determined by data availability. More

than 40% of all aquatic toxicity tests in the ECOTOX database (U.S. Environmental Protection Agency, 2019) report effects on mortality, making it the most frequently studied effect on aquatic organisms in this database. However, mortality is sometimes not the most important effect to consider, depending on the mode of action of the chemical under study. Additionally, the data used to derive standard endpoints (e.g. LC50 values) can be exploited further to obtain a more mechanistic understanding of sensitivity, for instance, by means of TKTD models.

Effects other than mortality might be ecologically more relevant, or more relevant due to the mode of action of the chemical. 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). Energy allocation has been suggested as a means to link 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 of sensitivity compared to when effects on mortality are used. More recently, energy allocation modelling has obtained renewed research interest under the acronym DEBtox (dynamic energy budget for toxicants), promoting simple generic models of animal life history (Baas et al., 2018; Jager et al., 2013; Kooijman, 2020).

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 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 et al., 2011; Rubach 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 under most standard test protocols (e.g. OECD, 2019). However, public access to these data remains difficult, either due to the requirements of journals where these studies are published, or, in case of regulatory studies, the rules of the regulatory frameworks. These difficulties can easily be overcome by a commitment to publish the raw data of experiments along with summary statistics like LC₅₀ values, preferably open access.

2.2. Exposure duration

Typically, acute toxicity tests with an exposure duration between 24 and 96 h are used for predictive modelling (Table 1). Again, this is primarily determined by data availability, since >50% of all aquatic toxicity test data available in the ECOTOX database concern tests with an exposure duration of up to 96 h (U.S. Environmental Protection Agency, 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 (fewer or smaller) effects after a 24 h continuous exposure than after a 96 h continuous 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

Table 1
Overview of modelling decisions made in the construction of interspecies correlation (IC), relatedness-based (RB), taxonomy-based (TB), and genomic-based (GB) models.^a

| | Effects (endpoint) | Exposure duration | Taxa | Transformation or normalization factor | 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, and aquatic invertebrates | logLC50 | µg/L | >1 chemical or MOA | Species | Linear regression | (e.g. Bejarano and Barron, 2014; Brill et al., 2016; Dyer et al., 2006; Feng et al., 2013) (Craig, 2013) |
| RB | Mortality (LC50), Immobilization (EC50) | 48 to 96 h | Fish, and aquatic invertebrates | logLC50 | µg/L | 1 chemical | Species | Bayesian regression | |
| | Mortality (LC50) | 96 h | Amphibians, fish, and aquatic invertebrates | logLC50 | µmol/L | > 1 MOA | Species | Bilinear regression | (Guénard et al., 2014) |
| | Mortality (LC50), Immobilization (EC50) | 24 to 96 h | Aquatic invertebrates | logLC50 | µg/L | > 1 heavy metal | Species | Bilinear regression | (Malaj et al., 2016) |
| | Population growth (EC50) | 96 h | Algae | $\sqrt{LC50}$ | µg/L | 1 chemical | Species | Multivariate analysis | (Larras 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 et al., 2010) |
| | Mortality (LC50), Immobilization (EC50) | 24 to 96 h | Aquatic invertebrates | $\log \left(\frac{1}{LC50} * F^b \right)$ | mol/L | 1 chemical | Genus, species | Genetic algorithm | (Ippolito et al., 2012) |
| | Mortality (LC50), Immobilization (EC50), Uptake (k_{in}), Elimination (k_{out}) | 48 h | Aquatic invertebrates | logLC50 | µg/L | 1 chemical | Species | Linear regression | (Rubach 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 et al., 2019) |
| GB | Mortality (LC50) | 48 and 96 h | Amphibians, fish, and aquatic invertebrates | logLC50 | µg/L | 1 chemical | Species | Linear regression | (LaLone et al., 2013) |
| | Mortality (LD50) | - | Birds | logLD50 | nmoles/kg | 3 chemicals | Species | Linear regression | (Farmahin et al., 2012) |

^a This table is intended to be illustrative, not exhaustive, due to space constraints.

^b Normalization factor was used to normalize the data according to exposure duration (Ippolito et al., 2012).

the chemical (e.g. Wiberg-Larsen 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 et al., 2014).

Besides running experiments long enough to ascertain that internal and external concentrations are in equilibrium, internal tissue concentrations could be measured and 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 et al., 2011). Focussing 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 et al., 2006).

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 (Gerhardt, 1993; Pascoe et al.,

1986). The biotic ligand model has been developed to examine the bio-availability 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.

There are many other variables that may be sources of variation in species sensitivity. Consider, for instance, the size (Poteat and Buchwalter, 2014), sex (McClellan-Green et al., 2007), and life stage (van der Lee et al., 2020) of the individuals used in the toxicity test. Although these sources of variation are well-known, setting additional selection criteria on them is nearly impossible, since reporting on these factors is not always, or has not always been, common practise under standard guidelines. Additionally, standard guidelines take a lot of time and effort to develop, and are therefore only available for a limited range of species, making the use of selection criteria on a wide range of species difficult. Similar as before, whether and which of these variables should be taken into consideration when determining toxicity depends on the compound and taxonomic group under study, since the importance of these variables depends on the combination of both. For instance, sex dependent responses towards endocrine disrupting compounds may be common among fish (Orlando and

Guillette, 2007), whilst they may be absent for certain groups of invertebrates due to the large complexity and variation in endocrine systems among species (Janer and Porte, 2007).

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 Section 2.5). Although $\mu\text{g l}^{-1}$ is still the most frequently used unit in aquatic toxicity tests (almost 50% of all aquatic tests available in the ECOTOX database, U.S. Environmental Protection Agency, 2019, and see Table 1), it is not the most suitable one. It is frequently overlooked that chemical sensitivity is primarily related to molecular activity, 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 (Escher and Hermens, 2002; Wezel and Opperhuizen, 1995), 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, U.S. Environmental Protection Agency, 2018) can help with converting mass units to molar units.

2.5. Grouping data, and its effects on explained variance

When data are limited, which is often the case, there is the possibility of grouping data (e.g. across chemicals or taxa) to obtain an adequately sized dataset suitable 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 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 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 et al., 1995).

Similar to using MOA to group across chemicals, higher taxonomic ranks (e.g. family, order) can be used to group across taxa, and may also be useful for reducing data gaps. Grouping at higher taxonomic ranks has the advantage of reducing bias due to extreme values and spurious data. However, potentially important differences in species sensitivity might be lost by summarising the sensitivity of several species at, for example, family level (Buchwalter et al., 2008; Ippolito 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 et al. (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. (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.

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 3.1), followed by a discussion of the merits and pitfalls associated with each of them (Section 3.2, Table 2), and close with a description on how the different predictor groups can be combined in a conceptual framework (Section 3.3).

3.1. Overview of methods

Interspecies-correlation (IC) models are log-linear least-squares regressions of the acute toxicity (E/LC_{50}) of chemicals measured in two species (e.g. Awkerman et al., 2008; Awkerman et al., 2014; Dyer et al., 2006; Dyer et al., 2008; Raimondo et al., 2007). 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 et al., 2016), aquatic invertebrates and vertebrates (e.g. Awkerman et al., 2014), terrestrial birds (e.g. Raimondo et al., 2007) and mammals (e.g. Awkerman et al., 2009), and have proven to be protective for rare and endangered species (Willming et al., 2016). However, not all predictions made by this kind of model 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 (e.g. belonging to the same order, Raimondo and Barron, 2019; Raimondo et al., 2007; Raimondo et al., 2010b).

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 et al., 2014; Malaj et al., 2016). The underlying principle of these models is that closely related species exhibit high correlation of sensitivity to chemicals, such that closely related species tend to have similar sensitivity, divergence of sensitivity, and uncertainty. These three aspects subsequently increase for more distantly related species. The correlation of the sensitivity of species with a known relatedness 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 higher taxonomic rank 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 et al., 2014; Malaj et al., 2016, Table 1).

Trait-based (TB) models use physiological, morphological and ecological characteristics of a species to describe its sensitivity towards chemical stressors (e.g. Rubach et al., 2010). Several traits of organisms are known to directly relate to organism sensitivity (e.g. larger organisms

Table 2
Brief description of the four groups of cross-species extrapolation approaches discussed in this review, along with information on their mechanistic explanation, data demand, and level of protection for ecological entities.

| | Main principle | Mechanistic explanation | Data demand | Protection of ecological entities |
|----|--|--|--|---|
| IC | Correlation between the responses of two species (surrogate and predicted species) to a range of chemicals | Absent | Toxicity data on multiple chemicals (both on the surrogate and predicted species) | Only the sensitivity of well-studied species can be predicted, and so far no examples of extrapolations to higher levels of biological organization exist. |
| RB | Evolutionary relatedness | Evolutionary related species exhibit similar sensitivity due to overlap in sensitivity-influencing traits and closely-related genetic patterns | Toxicity data, and data on taxonomic relatedness (i.e. a taxonomic or phylogenetic classification) | The sensitivity of real species assemblages can be predicted. Indirect effects of chemicals can only be predicted when chemical effects are restricted within taxonomic or phylogenetic groups carrying specific functions. |
| TB | Morphological, physiological and ecological relatedness | Differences in sensitivity-influencing morphological, physiological, or ecological characteristics of a species | Toxicity data, traits data, taxonomy data (to match toxicity and traits) | The sensitivity of real species assemblages can be predicted. Indirect effects of chemicals can be predicted based on what might happen to specific functional groups |
| GB | Similarity in biogeochemical pathways | Differences in sensitivity-influencing biogeochemical pathways | Toxicity data, adverse outcome pathway, data on one or more aspects of the biogeochemical pathway | Only the sensitivity of well-studied species can be predicted, and no examples exist yet on the extrapolation to higher levels of biological organization. |

tend to be more tolerant of toxicants) and therefore the relationships between these traits and sensitivity can be used to predict the sensitivity of untested species with known traits. Currently existing trait databases (e.g. [Usseglio-Polatera 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 et al., 2012](#); [Van den Berg et al., 2019](#)). Other traits that could help describe internal TD processes (e.g. presence of target receptors) are available, but have so far only been described for a small number of species (see Table 2 in [Rubach 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 ([Fedorenkova et al., 2010](#); [Snape et al., 2004](#)). Essentially, GB models directly link the genetic code underlying the molecules and pathways of chemical sensitivity to the sensitivity of the organism itself. Therefore, GB methods directly compare the differences between how organisms respond to chemicals internally, rather than the extent of relatedness in RB methods or the traits (which may have multiple genetic or phenotypic origins) of TB models that both partially relate to organism sensitivity. 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 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 et al., 2001](#); [Poynton 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, for which adverse outcome pathways (AOPs) have been suggested as a suitable framework ([Ankley 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 1).

3.2. Comparison of methods

3.2.1. Mechanistic explanation

[Raimondo et al. \(2010a\)](#) 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 chemical untested on the target species 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 et al., 2014](#); [Malaj 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 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 focusing on aquatic invertebrates has, for instance, demonstrated that the uptake rate of chemicals can to a large extent be explained by the lipid content of an organism, whilst elimination rates are negatively correlated with the degree of sclerotization ([Rubach 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 Table 2 in [Rubach et al. \(2011\)](#) for an overview of the availability of a wide range of traits for algae, fish, aquatic plants, birds, mammals, and aquatic invertebrates, and the strength of the trait-process relationship (i.e. plausible but not proven, some evidence for some taxa, relationship available for several taxa).

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 Forbes et al., 2006 for an overview of the limitations of biomarkers for assessing population level effects). Even if a complete AOP is available, capturing all possible molecular initiating events and/or key events that could be generated by the compound under study, 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 et al. (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 et al., 2013).

3.2.2. Data demand

IC models only require data on toxicity (e.g. EC_{50} , LC_{50}), which can be obtained from public databases such as the ECOTOX Knowledgebase (U.S. Environmental Protection Agency, 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 et al., 2010a). Nevertheless, the latest IC models for aquatic animals contain >8500 toxicity values covering 316 species and 1499 chemicals (Raimondo et al., 2015). However, the taxonomic coverage of these models is restricted, with >60% of all the models available in WebICE extrapolating from one fish species to another (Raimondo et al., 2015), and of another 26%, 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 et al., 2014), and large projects are available to sequence vertebrate genomes (e.g. the Genome 10 K project, Koepfli 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, 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 literature (see Table 2 in Rubach et al., 2011). The study of Van den Berg et al. (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 et al., 2000), whilst more trait databases are available for invertebrates (Hébert et al., 2016; Poff et al., 2006; Schäfer et al., 2011). For fish, a wide range of traits are available, distributed over several trait databases (Frimpong and Angermeier, 2009; Froese and Pauly, 2000;

Lamouroux et al., 2002) and covering a large part of the taxonomic diversity of fish. For algae we are aware of two traits databases currently available (Lange et al., 2016; Reynolds et al., 2002), 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. The taxonomic nomenclature used in the traits database has to exactly match the one used in the toxicity database. If this is not the case, the taxonomy of both the traits and the toxicity database has to be standardized by means of an external taxonomy database. Access to taxonomic data has already been described under RB models.

GB models are the most data demanding, because they require peer-reviewed AOPs, based on validated biomarkers. Currently, 274 AOPs have been described in the AOP wiki in total covering 521 stressors (including chemicals, environmental factors), although the OECD status of the majority of them remains 'under development' (<https://aopwiki.org/>, accessed on the 25th of January 2020), 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 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 et al., 2019) or enzymatic markers (Arini 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, Fig. 3 in Zhang et al., 2018). However, these methods are time-, and cost-intensive, and are frequently incomparable due to inconsistent bioinformatic methods for data filtering, concentration-response modelling and quantitative characterization of genes and pathways (Zhang et al., 2018).

3.2.3. 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 ERA by filling in data gaps. However, the methods studied in this review vary in two important ways: i) in the way they are able to 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.

Researchers have known for a long time that real species assemblages vary through time (Murphy, 1978) and space (Vannote 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, provided that data requirements of the species whose sensitivity you want to predict are available or can be collected. This contrasts with IC models, which require sufficient toxicity data to be available for the taxon whose sensitivity we want to predict (Section 3.2.2), and then still might be overfitted to the training data due to the absence of mechanistic relationships. GB models require, at least, to have the part of the genome sequenced that is associated with the key molecular initiating event(s) (LaLone et al., 2013). This is to ensure that divergence of genomic sequences linked to the molecular targets of a chemical can be associated with differences in the sensitivity between species. Consequently, extensive collection of genomic data and understanding of the chemical's toxicity pathway is required to produce a robust GB model. Therefore, IC and GB models are only able to predict the sensitivity of well-studied 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 an 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. 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 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 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 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 algae, 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 be used for the extrapolation of 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 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 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 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.

3.3. A combined approach 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 a conceptual framework. Indeed, all studies discussed in this review (Table 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 et al., 2012; Poteat et al., 2015).

That 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 (Fig. 2). Studies describing species differences in TK parameters (e.g. Buchwalter et al., 2008; Rubach et al., 2012) found that traits like mode of respiration, body size and other morphological traits 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 of differences in the TD part of species sensitivity (e.g. as found in Larras et al., 2014). So we can hypothesise that TB approaches are good in explaining the TK part of differences in species sensitivity, whilst GB approaches are good in explaining the TD part of differences in species sensitivity (Fig. 2). Additionally, RB approaches have the potential to represent aspects of both TK and TD processes, 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). Therefore, RB predictors can be added to the model to represent sensitivity related processes that are still unknown (Fig. 2). Alternatively, a stand-alone RB analysis can be used to distinguish which taxa are sensitive and tolerant to a specific chemical or MOA. This information can help ease the search for molecular target(s) or traits powerful in describing differences in species sensitivity, since it must be due to genomic or trait differences existing between sensitive and tolerant taxa. Finally, IC models can be used if the MOA of the chemical under study has been extensively studied before, and if the taxonomic coverage of these models is sufficient to determine the potential risk to non-target organisms.

Considering that the best performing models can be found by combining the different methods in a conceptual framework, the different layers (IC, RB, TB, GB) of the TK and TD processes as illustrated in Fig. 2 can be regarded as different levels of a tiered approach, each level introducing more complexity and mechanistic explanation. At the lowest level of this approach, you can find IC models, which can be used for a preliminary hazard assessment. For this, existing IC models should be collected and applied to conduct a preliminary assessment of hazard following a weight-of-evidence approach. Besides evaluating the potential risk to non-target species, the used models should be assessed on their taxonomic coverage and model performance, whose thresholds should be set beforehand. The thresholds of the taxonomic coverage and model performance will depend on the trade-off between the purpose of the modelling effort (i.e. to support priority setting procedures, to supplement the use of experimental data in weight-of-evidence approaches, or to completely substitute the need for experimental data) and the strictness of the regulatory framework that the target compound falls under (some being more conservative than others). At the end of every tier, an evaluation is done to check whether the risks are shown to be negligible or acceptable with reasonable certainty, and whether enough information is available to make a regulatory decision. If the evaluation still indicates a potential risk to certain non-target organisms or further information is required for decision making, continuation to the next tier is necessary.

In the higher levels of this approach, predictor groups are added according to their data availability. First, the most abundantly available and easily accessible data is added to the models: taxonomic relatedness. Model construction is done anew, followed by an evaluation of the risks, taxonomic coverage, and model performance. If necessary,

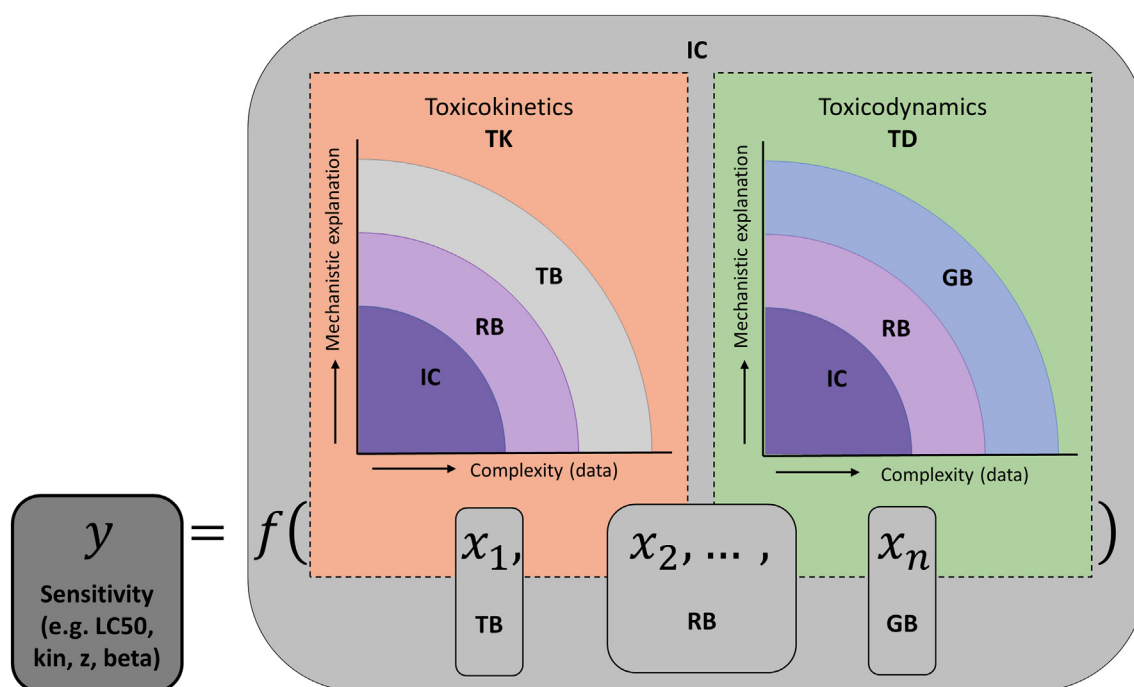


Fig. 2. An abstract visualization of the conceptual framework suggested to combine the different modelling approaches (IC, RB, TB and GB) discussed in this review. The different layers (IC, RB, TB, GB) of the TK and TD processes can be regarded as the steps of a tiered approach, increasing in complexity and mechanistic explanation.

we continue to the next level, in which trait predictors are introduced. For this, a hypothesis-driven approach is used to select sensitivity-related traits. In the case that sensitivity-related traits of the taxocompound combination are unknown, the previous RB approach can be used to focus research. For instance, the RB approach has distinguished certain taxonomic groups as sensitive or tolerant. A study of the traits belonging with these taxonomic groups can assist in creating hypotheses regarding sensitivity-related traits. If traits data are insufficiently available in existing databases, new traits data can be collected using literature research or measuring the traits in the laboratory. Once sufficient traits data are available, TB-RB models can be constructed, and risk and model evaluation is repeated. In the next and final level of this approach, more mechanistic information can be added to the models by introducing GB predictors. For this, molecular markers important for the MOA of the target compound under study need to be known and available. If this is not the case, the RB approach can be used to focus research, similarly as how this was done for traits. Once sufficient data are available, TB-GB models can be constructed, potentially supplemented with RB predictors to represent any missing molecular markers or traits that are important for describing the sensitivity process. Only when it is still not clear whether the risk conclusion is acceptable after the final risk and model evaluation, execution of experiments following one of the more traditional tiered approaches is necessary.

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 modellers 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 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 outcome.

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 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 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 et al., 2013). Research performed on a GB based approach studied the influence of different preliminary variable selection methods on model performance (Mannheimer 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 might be the case for GB approaches. For smaller datasets, however, preliminary variable selection methods can have a severe impact on the modelling results. Predictors should in that case 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 et al., 2001; Schultz and Cronin, 2003). Traits like size and number of offspring per clutch are known to vary over space (Orlofske and Baird, 2014), and are additionally recognized to alter ecological dynamics through indirect effects (Bolnick et al.,

2011). Therefore, the more predictors included in the model, the larger the chance of incorporating errors. Extrapolating the variation associated with predictors 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 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 et al., 2016; Moermond 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 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. Ladrone 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 et al., 2019). Whether a linear or non-linear method is more suitable depends on the hypothesised relationship between the dependent and independent variables, the number of independent variables available, and on the degree of mechanistic information contained within these independent variables. 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. Rico and Van den Brink, 2015; Rubach et al., 2012; Rubach et al., 2010). 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. Another crucial approach to avoid overfitting is to perform a model validation step. This can be done by splitting the data in a training and a test set. The model is then fitted to the training data, before being evaluated on the test data. In this way, the model can be evaluated on its predictive power, rather than on its fit. Doing this in a repeated, randomized manner is called cross-validation. However, it is important to realize that a (cross-)validation exercise is primarily feasible when the dataset is sufficiently large. When data are limited, bad validation results do not necessarily indicate an erroneous relationship, and literature might be available to provide support for the found relationship. However, good validation results provide proof that the found relationship is consistent among the available data, and that the model is not performing well merely due to coincidence.

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 and performance of these models does to a large extent depend on the modelling 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 reimplementation multiple times. In this way, we can spend our efforts on using and elaborating on existing work.

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 (Fig. 2). For these reasons, we suggest a conceptual framework (Fig. 2), combining predictors describing important traits determining the uptake and elimination of chemicals (e.g. size, respiration mode, exoskeleton-thickness), with the amount of sequence similarity in molecular targets, and relatedness predictors utilised where data for traits and molecular targets are unavailable. This conceptual framework can be considered a tiered approach, where moving up a tier equals moving up in level of complexity and mechanistic understanding of the sensitivity process. We realize that the conceptual framework suggested in Section 3.3 needs to be developed further to enable practical application in regulatory risk assessment. A more detailed, set-by-step framework, supplemented with case studies demonstrating potential practical applications, will be of great importance for moving this field forward.

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.

So, after answering these three sub-questions, is it now clear how to extrapolate chemical sensitivity across species? For some of the methods discussed in this review, this is indeed straight forward, and in some occasions they have already been used in regulatory risk assessment. For instance, IC models matching model requirements can directly be used in regulatory risk assessment. However, for cross-species extrapolation methods to really find its way into regulatory risk assessment, additional work will have to be done, especially in the area of their uncertainty and

practical applicability. As briefly has been mentioned before in Section 3.3, the requirements of the modelling effort (e.g. acceptable uncertainty boundaries) will depend on the trade-off between the purpose of the modelling effort (i.e. to support priority setting procedures, to supplement the use of experimental data in weight-of-evidence approaches, or to completely substitute the need for experimental data) and the strictness of the regulatory framework that the target compound falls under (some being more conservative than others). For example, when models are applied to support priority setting, or to supplement experimental data in weight-of-evidence approaches, their use is more indirect. Under these circumstances, experimental data and other information is available, making the extrapolation results not likely to be decisive in the final assessment. However, when the objective is to replace 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. Especially in the latter case, a firm grip on the uncertainty associated with these models is necessary. Without concrete measures of uncertainty, modelling outcomes will have to be supplemented with something similar to the assessment factors that we considered unspecific and therefore inappropriate for risk assessment purposes.

Considering additional work on the practical applicability of cross-species extrapolation models, the main focus should lie on developing the conceptual framework suggested here in more detail. Working through some case studies will demonstrate how feasible the suggested approach is, and which research fields will need to evolve more before practical implementation becomes possible. For example, which difficulties lie in the application of RB and TB methods to still unknown taxonomic- or trait profiles? Will they indeed be able to accurately predict the sensitivity of natural species assemblages, or will their species coverage remain too low? Considering GB approaches, however promising they sound, will it really become possible to use approaches like this for a wide range of species, or will we get lost in the maze of AOPs, genetic markers, and key events? Finally, the question remains whether the current surge for open science and reproducible research will really turn the field of ecotoxicology into ART (accurate, reliable, and transparent), or that crucial data and information will remain hidden behind walls of journal requirements and regulatory frameworks? It is only after these things become clear, that we will know how we can extrapolate species sensitivity. This would offer opportunities for refining risk assessments, including spatial and temporal consideration of sensitivity, and provide methods for reducing animal testing and the costs associated with them.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Unilever under the title 'Ecological scenarios and models for use in risk assessment' (ESMU), by CEFIC under the GETREAL project, by the University of Sheffield Prize Scholarship, and by the Natural Environment Research Council (NERC). There are no conflicts of interest. Thanks to Lara Schuijt for the many discussions and feedbacks.

References

Amiard-Triquet, C., 2009. Behavioral disturbances: the missing link between sub-organismal and supra-organismal responses to stress? Prospects based on aquatic research. *Hum. Ecol. Risk Assess.* 15, 87–110.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., et al., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29, 730–741.

Arini, A., Mittal, K., Dornbos, P., Head, J., Rutkiewicz, J., Basu, N., 2017. A cell-free testing platform to screen chemicals of potential neurotoxic concern across twenty vertebrate species. *Environ. Toxicol. Chem.* 36, 3081–3090.

Asfaw, A., Ellersieck, M.R., Mayer, F.L., 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.

Awkerman, J.A., Raimondo, S., Barron, M.G., 2008. Development of species sensitivity distributions for wildlife using interspecies toxicity correlation models. *Environ. Sci. Technol.* 42, 3447–3452.

Awkerman, J.A., Raimondo, S., Barron, M.G., 2009. Estimation of wildlife hazard levels using interspecies correlation models and standard laboratory rodent toxicity data. *J. Toxic. Environ. Health A* 72, 1604–1609.

Awkerman, J.A., Raimondo, S., Jackson, C.R., Barron, M.G., 2014. Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models. *Environ. Toxicol. Chem.* 33, 688–695.

Baas, J., Augustine, S., Marques, G.M., Dorne, J.-L., 2018. Dynamic energy budget models in ecological risk assessment: from principles to applications. *Sci. Total Environ.* 628, 249–260.

Bejarano, A.C., Barron, M.G., 2014. Development and practical application of petroleum and dispersant interspecies correlation models for aquatic species. *Environ. Sci. Technol.* 48, 4564–4572.

Benfenati, E., Piclin, N., Roncaglioni, A., Vari, M., 2001. Factors influencing predictive models for toxicology. *SAR QSAR Environ. Res.* 12, 593–603.

Biggs, J., Williams, P., Whitfield, M., Nicolet, P., Brown, C., Hollis, J., et al., 2007. The freshwater biota of British agricultural landscapes and their sensitivity to pesticides. *Agric. Ecosyst. Environ.* 122, 137–148.

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

Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., et al., 2011. Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* 26, 183–192.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al., 2001. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* 29, 365.

Brill, J.L., Belanger, S.E., Chaney, J.G., Dyer, S.D., Raimondo, S., Barron, M.G., et al., 2016. Development of algal interspecies correlation estimation models for chemical hazard assessment. *Environ. Toxicol. Chem.* 35, 2368–2378.

Brock, T.C.M., Arts, G.H.P., Maltby, L., Van den Brink, P.J., 2006. Aquatic risks of pesticides, ecological protection goals, and common aims in European Union legislation. *Integr. Environ. Assess. Manag.* 2, e20–e46.

Brown, A.R., Whale, G., Jackson, M., Marshall, S., Hamer, M., Solga, A., et al., 2017. Toward the definition of specific protection goals for the environmental risk assessment of chemicals: a perspective on environmental regulation in Europe. *Integr. Environ. Assess. Manag.* 13, 17–37.

Buchwalter, D.B., Cain, D.J., Martin, C.A., Xie, L., Luoma, S.N., Garland, T., 2008. Aquatic insect ecophysiological traits reveal phylogenetically based differences in dissolved cadmium susceptibility. *Proc. Natl. Acad. Sci.* 105, 8321–8326.

Calow, P., Sibly, R., 1990. A physiological basis of population processes: ecotoxicological implications. *Funct. Ecol.* 4, 283–288.

Carriger, J.F., Martin, T.M., Barron, M.G., 2016. A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors. *Aquat. Toxicol.* 180, 11–24.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., et al., 2014. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57, 4977–5010.

Clements, W.H., Rohr, J.R., 2009. Community responses to contaminants: using basic ecological principles to predict ecotoxicological effects. *Environ. Toxicol. Chem.* 28, 1789–1800.

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.

De Coen, W.M., Janssen, C.R., 2003. The missing biomarker link: relationships between effects on the cellular energy allocation biomarker of toxicant-stressed *Daphnia magna* and corresponding population characteristics. *Environ. Toxicol. Chem.* 22, 1632–1641.

De Laender, F., Morselli, M., Baveco, H., Van den Brink, P.J., Di Guardo, A., 2015. Theoretically exploring direct and indirect chemical effects across ecological and exposure scenarios using mechanistic fate and effects modelling. *Environ. Int.* 74, 181–190.

del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., Ruiz-Trillo, I., 2014. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* 29, 252–259.

Donkin, P., 2009. Quantitative structure-activity relationships. *Handbook of Ecotoxicology*. Blackwell Publishing Ltd., pp. 785–811.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46.

Dyer, S.D., Versteeg, D.J., Belanger, S.E., Chaney, J.G., Mayer, F.L., 2006. Interspecies correlation estimates predict protective environmental concentrations. *Environ. Sci. Technol.* 40, 3102–3111.

Dyer, S.D., Versteeg, D.J., Belanger, S.E., Chaney, J.G., Raimondo, S., Barron, M.G., 2008. Comparison of species sensitivity distributions derived from interspecies correlation models to distributions used to derive water quality criteria. *Environ. Sci. Technol.* 42, 3076–3083.

EFSA, 2014. Scientific opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. *EFSA J.* 12, 3589.

EFSA PPR Panel (Panel on Plant Protection Products and their Residues), Ockleford C., Adriaanse P., Bery P, Bro, ck, T., Duquesne, S., et al., 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 J.* 16, e05377.
- Eisner, B., Doering, J., Beitel, S., Hecker, M., 2019. Cross-species Extrapolation Using a Simplified *In Vitro* Tissue Explant Assay in Fish.
- Erickson, R.J., 2013. The biotic ligand model approach for addressing effects of exposure water chemistry on aquatic toxicity of metals: genesis and challenges. *Environ. Toxicol. Chem.* 32, 1212–1214.
- Escher, B.I., Hermens, J.L.M., 2002. Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. *Environ. Sci. Technol.* 36, 4201–4217.
- 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* 2000; 22.
- Farmahin, R., Wu, D., Crump, D., Hervé, J.C., Jones, S.P., Hahn, M.E., et al., 2012. Sequence and *in vitro* function of chicken, ring-necked pheasant, and Japanese quail AHR1 predict *in vivo* sensitivity to dioxins. *Environ. Sci. Technol.* 46, 2967–2975.
- Federhen, S., 2011. The NCBI Taxonomy database. *Nucleic Acids Res.* 40, D136–D143.
- Fedorenkova, A., Vonk, J.A., Lenders, H.R., Ouborg, N.J., Breure, A.M., Hendriks, A.J., 2010. Ecotoxicogenomics: bridging the gap between genes and populations. *Environ. Sci. Technol.* 44, 4328–4333.
- Feng, C., Wu, F., Dyer, S., Chang, H., Zhao, X., 2013. Derivation of freshwater quality criteria for zinc using interspecies correlation estimation models to protect aquatic life in China. *Chemosphere* 90, 1177–1183.
- Forbes, V.E., Palmqvist, A., Bach, L., 2006. The use and misuse of biomarkers in ecotoxicology. *Environ. Toxicol. Chem.* 25, 272–280.
- Friant, S., Henry, L., 1985. Relationship between toxicity of certain organic compounds and their concentrations in tissues of aquatic organisms: a perspective. *Chemosphere* 14, 1897–1907.
- Frimpong, E.A., Angermeier, P.L., 2009. Fish traits: a database of ecological and life-history traits of freshwater fishes of the United States. *Fisheries* 34, 487–495.
- Froese, R., Pauly, D., 2000. *FishBase 2000: Concepts Designs and Data Sources.* vol 1594. WorldFish.
- Gerhardt, A., 1993. Review of impact of heavy metals on stream invertebrates with special emphasis on acid conditions. *Water Air Soil Pollut.* 66, 289–314.
- Gleason, T.R., Nacci, D.E., 2001. Risks of endocrine-disrupting compounds to wildlife: extrapolating from effects on individuals to population response. *Hum. Ecol. Risk Assess.* 7, 1027–1042.
- Guénard, G., von der Ohe, P.C., Walker, S.C., Lek, S., Legendre, P., 2014. Using phylogenetic information and chemical properties to predict species tolerances to pesticides. *Proc. R. Soc. Lond. B Biol. Sci.* 281, 20132329.
- Hammers-Wirtz, M., Ratte, H.T., 2000. Offspring fitness in *Daphnia*: is the *Daphnia* reproduction test appropriate for extrapolating effects on the population level? *Environ. Toxicol. Chem.* 19, 1856–1866.
- Hébert, M.-P., Beisner, B.E., Maranger, R., 2016. A compilation of quantitative functional traits for marine and freshwater crustacean zooplankton. *Ecology* 97, 1081.
- Hendriks, A.J., van der Linde, A., Cornelissen, G., Sijm, D.T., 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. *Environ. Toxicol. Chem.* 20, 1399–1420.
- Hickey, C.W., Clements, W.H., 1998. Effects of heavy metals on benthic macroinvertebrate communities in New Zealand streams. *Environ. Toxicol. Chem.* 17, 2338–2346.
- Ippolito, A., Todeschini, R., Vighi, M., 2012. Sensitivity assessment of freshwater macroinvertebrates to pesticides using biological traits. *Ecotoxicology* 21, 336–352.
- ITIS, 2019. *Integrated Taxonomic Information System.*
- Jager, T., Heugens, E.H., Kooijman, S.A., 2006. Making sense of ecotoxicological test results: towards application of process-based models. *Ecotoxicology* 15, 305–314.
- Jager, T., Albert, C., Preuss, T.G., Ashauer, R., 2011. General unified threshold model of survival-a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environ. Sci. Technol.* 45, 2529–2540.
- Jager, T., Martin, B.T., Zimmer, E.I., 2013. DEBkiss or the quest for the simplest generic model of animal life history. *J. Theor. Biol.* 328, 9–18.
- Janardan, S.K., Olson, C.S., Schaeffer, D.J., 1984. Quantitative comparisons of acute toxicity of organic chemicals to rat and fish. *Ecotoxicol. Environ. Saf.* 8, 531–539.
- Janer, G., Porte, C., 2007. Sex steroids and potential mechanisms of non-genomic endocrine disruption in invertebrates. *Ecotoxicology* 16, 145–160.
- Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108.
- Kase, R., Korkaric, M., Werner, I., Ågerstrand, M., 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. *Environ. Sci. Eur.* 28, 7.
- Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., Embry, M.R., 2017. Mode of action (MOA) assignment classifications for ecotoxicology: an evaluation of approaches. *Environ. Sci. Technol.* 51, 10203–10211.
- Klaper, R., Thomas, M.A., 2004. At the crossroads of genomics and ecology: the promise of a canary on a chip. *BioScience* 54, 403–412.
- Koepfli, K.-P., Paten, B., Genome, KCoS, O'Brien, S.J., 2015. The genome 10K project: a way forward. *Annu. Rev. Anim. Biosci.* 3, 57–111.
- Kooijman, S., 1987. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Res.* 21, 269–276.
- Kooijman, S.A., 2020. The standard dynamic energy budget model has no plausible alternatives. *Ecol. Model.* 428, 109106.
- Ladroue, C., Guo, S., Kendrick, K., Feng, J., 2009. Beyond element-wise interactions: identifying complex interactions in biological processes. *PLoS One* 4.
- LaLone, C.A., Villeneuve, D.L., Burgoon, L.D., Russom, C.L., Helgen, H.W., Berninger, J.P., et al., 2013. Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action. *Aquat. Toxicol.* 144–145, 141–154.
- Lamouroux, N., Poff, N.L., Angermeier, P.L., 2002. Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients. *Ecology* 83, 1792–1807.
- Lange, K., Townsend, C.R., Matthaei, C.D., 2016. A trait-based framework for stream algal communities. *Ecol. Evol.* 6, 23–36.
- Larras, F., Keck, F., Montuelle, B., Rimet, F., Bouchez, A.S., 2014. Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? *Environ. Sci. Technol.* 48, 1921–1930.
- Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., et al., 2018. Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci.* 115, 4325–4333.
- Malaj, E., Guénard, G., Schäfer, R.B., Ohe, P.C., 2016. Evolutionary patterns and physico-chemical properties explain macroinvertebrate sensitivity to heavy metals. *Ecol. Appl.* 26, 1249–1259.
- Mannheimer, J.D., Duval, D.L., Prasad, A., Gustafson, D.L., 2019. A systematic analysis of genomics-based modeling approaches for prediction of drug response to cytotoxic chemotherapies. *BMC Med. Genet.* 12, 87.
- Mayer, F.L., Ellersieck, M.R., 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., Landrum, P., Luoma, S., Meador, J., Merten, A., Shephard, B., et al., 2011. Advancing environmental toxicology through chemical dosimetry: external exposures versus tissue residues. *Integr. Environ. Assess. Manag.* 7, 7–27.
- McClellan-Green, P., Romano, J., Oberdörster, E., 2007. Does gender really matter in contaminant exposure? A case study using invertebrate models. *Environ. Res.* 104, 183–191.
- Moermond, C.T., Kase, R., Korkaric, M., Ågerstrand, M., 2016. CRED: criteria for reporting and evaluating ecotoxicity data. *Environ. Toxicol. Chem.* 35, 1297–1309.
- Murphy, P., 1978. The temporal variability in biotic indices. *Environ. Pollut.* (1970) 17, 227–236.
- Nendza, M., Muller, M.T., 2000. Discriminating toxicant classes by MOA: 2. Physico-chemical descriptors. *Quantitative Structure-Activity Relationships.* vol. 19, pp. 581–598.
- OECD, 2019. *Test No. 203: Fish, Acute Toxicity Test.*
- Orlando, E.F., Guillelte, L.J., 2007. Sexual dimorphic responses in wildlife exposed to endocrine disrupting chemicals. *Environ. Res.* 104, 163–173.
- Orlofske, J.M., Baird, D.J., 2014. Incorporating continuous trait variation into biomonitoring assessments by measuring and assigning trait values to individuals or taxa. *Freshw. Biol.* 59, 477–490.
- Pascoe, D., Evans, S.A., Woodworth, J., 1986. Heavy metal toxicity to fish and the influence of water hardness. *Arch. Environ. Contam. Toxicol.* 15, 481–487.
- Pennie, B., Ahr, H.-J., Braun, C., Bars, R., Jacob, E., Carpanini, F., 2001. *Genomics, Transcript Profiling, Proteomics and Metabonomics (GTPM) an Introduction (White Paper).* ECETOC, Brussels.
- Poff, N.L., Olden, J.D., Vieira, N.K., Finn, D.S., Simmons, M.P., Kondratieff, B.C., 2006. Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. *J. N. Am. Benthol. Soc.* 25, 730–755.
- Poteat, M.D., Buchwalter, D.B., 2014. Phylogeny and size differentially influence dissolved Cd and Zn bioaccumulation parameters among closely related aquatic insects. *Environ. Sci. Technol.* 48, 5274–5281.
- Poteat, M.D., Jacobus, L.M., Buchwalter, D.B., 2015. The importance of retaining a phylogenetic perspective in traits-based community analyses. *Freshw. Biol.* 60, 1330–1339.
- Poynton, H.C., Robinson, W.E., Blalock, B.J., Hannigan, R.E., 2014. Correlation of transcriptomic responses and metal bioaccumulation in *Mytilus edulis* L. reveals early indicators of stress. *Aquat. Toxicol.* 155, 129–141.
- Raimondo, S., Barron, M., 2019. Application of Interspecies Correlation Estimation (ICE) models and QSAR in estimating species sensitivity to pesticides. *SAR QSAR Environ. Res.* 1–18.
- Raimondo, S., Mineau, P., Barron, M.G., 2007. Estimation of chemical toxicity to wildlife species using interspecies correlation models. *Environ. Sci. Technol.* 41, 5888–5894.
- Raimondo, S., Jackson, C.R., Barron, M.G., 2010a. Influence of taxonomic relatedness and chemical mode of action in acute interspecies estimation models for aquatic species. *Environ. Sci. Technol.* 44, 7711–7716.
- Raimondo, S., Vivian, D.N., Barron, M.G., 2010b. 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.
- Raimondo, S., Lilavois, C.R., Barron, M.G., 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.
- Reynolds, C.S., Huszar, V., Kruk, C., Naselli-Flores, L., Melo, S., 2002. Towards a functional classification of the freshwater phytoplankton. *J. Plankton Res.* 24, 417–428.
- Rico, A., Van den Brink, P.J., 2015. Evaluating aquatic invertebrate vulnerability to insecticides based on intrinsic sensitivity, biological traits, and toxic mode of action. *Environ. Toxicol. Chem.* 34, 1907–1917.
- Rosenfeld, J.S., 2002. Functional redundancy in ecology and conservation. *Oikos* 98, 156–162.
- Rubach, M.N., Baird, D.J., Van den Brink, P.J., 2010. A new method for ranking mode-specific sensitivity of freshwater arthropods to insecticides and its relationship to biological traits. *Environ. Toxicol. Chem.* 29, 476–487.
- Rubach, M.N., Ashauer, R., Buchwalter, D.B., De Lange, H.J., Hamer, M., Preuss, T.G., et al., 2011. Framework for traits-based assessment in ecotoxicology. *Integr. Environ. Assess. Manag.* 7, 172–186.

- Rubach, M.N., Baird, D.J., Boerwinkel, M.-C., Maund, S.J., Roessink, I., Van den Brink, P.J., 2012. Species traits as predictors for intrinsic sensitivity of aquatic invertebrates to the insecticide chlorpyrifos. *Ecotoxicology* 21, 2088–2101.
- Schäfer, R.B., Kefford, B.J., Metzeling, L., Liess, M., Burgert, S., Marchant, R., et al., 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. *Sci. Total Environ.* 409, 2055–2063.
- Schultz, T.W., Cronin, M.T., 2003. Essential and desirable characteristics of ecotoxicity quantitative structure–activity relationships. *Environ. Toxicol. Chem.* 22, 599–607.
- Segner, H., 2011. Chapter 86 - reproductive and developmental toxicity in fishes. In: Gupta, R.C. (Ed.), *Reproductive and Developmental Toxicology*. Academic Press, San Diego, pp. 1145–1166.
- Snape, J.R., Maund, S.J., Pickford, D.B., Hutchinson, T.H., 2004. Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquat. Toxicol.* 67, 143–154.
- Society of Environmental Toxicology and Chemistry, 2019. *Technical Issue Paper: Recommended Minimum Reporting Information for Environmental Toxicity Studies*.
- Suter, G.W., 2016. *Ecological Risk Assessment*. CRC press.
- U.S. Environmental Protection Agency, 2018. *Estimation Programs Interface Suite for Microsoft Windows*. Agency USEP, Washington, DC, USA.
- U.S. Environmental Protection Agency, 2019. *ECOTOX user guide: ECOTOXicology knowledgebase system*. U.S. Environmental Protection Agency <http://www.epa.gov/ecotox/>.
- Usseglio-Polatera, P., Bournaud, M., Richoux, P., Tachet, H., 2000. Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits. *Freshw. Biol.* 43, 175–205.
- Van den Berg, S.J.P., Baveco, H., Butler, E., De Laender, F., Focks, A., Franco, A., et al., 2019. Modeling the sensitivity of aquatic macroinvertebrates to chemicals using traits. *Environ. Sci. Technol.* 53, 6025–6034.
- Van den Brink, P.J., 2008. Ecological risk assessment: from book-keeping to chemical stress ecology. *Environ. Sci. Technol.* 42, 8999–9004.
- Van den Brink, P.J., Alexander, A.C., Desrosiers, M., Goedkoop, W., Goethals, P.L., Liess, M., et al., 2011. Traits-based approaches in bioassessment and ecological risk assessment: strengths, weaknesses, opportunities and threats. *Integr. Environ. Assess. Manag.* 7, 198–208.
- van der Lee, G.H., Kraak, M.H.S., Verdonschot, R.C.M., Verdonschot, P.F.M., 2020. Persist or perish: critical life stages determine the sensitivity of invertebrates to disturbances. *Aquat. Sci.* 82, 24.
- van Wezel, A.P., de Vries, D.A.M., Kostense, S., Sijm, D.T.H.M., Opperhuizen, A., 1995. Intra-species variation in lethal body burdens of narcotic compounds. *Aquat. Toxicol.* 33, 325–342.
- Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. *Can. J. Fish. Aquat. Sci.* 37, 130–137.
- Wezel, A.P., Opperhuizen, A., 1995. Narcosis due to environmental pollutants in aquatic organisms: residue-based toxicity, mechanisms, and membrane burdens. *Crit. Rev. Toxicol.* 25, 255–279.
- Wiberg-Larsen, P., Graeber, D., Kristensen, E.A., Baatrup-Pedersen, A., Friberg, N., Rasmussen, J.J., 2016. Trait characteristics determine pyrethroid sensitivity in non-standard test species of freshwater macroinvertebrates: a reality check. *Environ. Sci. Technol.* 50, 4971–4978.
- Willming, M.M., Lilavois, C.R., Barron, M.G., Raimondo, S., 2016. Acute toxicity prediction to threatened and endangered species using Interspecies Correlation Estimation (ICE) models. *Environ. Sci. Technol.* 50, 10700–10707.
- Wintle, B.A., McCarthy, M.A., Volinsky, C.T., Kavanagh, R.P., 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conserv. Biol.* 17, 1579–1590.
- Yang, J., Zhang, X., Xie, Y., Song, C., Sun, J., Zhang, Y., et al., 2017. Ecogenomics of zooplankton community reveals ecological threshold of Ammonia nitrogen. *Environ. Sci. Technol.* 51, 3057–3064.
- Zhang, X., Xia, P., Wang, P., Yang, J., Baird, D.J., 2018. *Omic Advances in Ecotoxicology*. ACS Publications.