Modeling Regulatory Threshold Levels for Pesticides in Surface Waters from Effect Databases

Lara L. Petschick [1], **Sascha Bub [1]**, **Jakob Wolfram [1]**, **Sebastian Stehle [1,2]** and **Ralf Schulz [1,*]**

[1]   iES Landau, Institute for Environmental Sciences, University of Koblenz-Landau,
      D-76829 Landau, Germany; petschick@uni-landau.de (L.L.P.); bub@uni-landau.de (S.B.);
      wolfram@uni-landau.de (J.W.); stehle@uni-landau.de (S.S.)
[2]   Eusserthal Ecosystem Research Station, University of Koblenz-Landau, D-76857 Eusserthal, Germany
[*]   Correspondence: schulz@uni-landau.de

Abstract: Regulatory threshold levels (RTL) represent robust benchmarks for assessing risks of pesticides, e.g., in surface waters. However, comprehensive scientific risk evaluations comparing RTL to measured environmental concentrations (MEC) of pesticides in surface waters were yet restricted to a low number of pesticides, as RTL are only available after extensive review of regulatory documents. Thus, the aim of the present study was to model RTL equivalents (RTLe) for aquatic organisms from publicly accessible ecotoxicological effect databases. We developed a model that applies validity criteria in accordance with official US EPA review guidelines and validated the model against a set of manually retrieved RTL (n = 49). Model application yielded 1283 RTLe (n = 676 for pesticides, plus 607 additional RTLe for other use types). In a case study, the usability of RTLe was demonstrated for a set of 27 insecticides by comparing RTLe and RTL exceedance rates for 3001 MEC from US surface waters. The provided dataset enables thorough risk assessments of surface water exposure data for a comprehensive number of substances. Especially regions without established pesticide regulations may benefit from this dataset by using it as a baseline information for pesticide risk assessment and for the identification of priority substances or potential high-risk regions.Dataset License: CC-BY-SA

Dataset: The dataset can be found in Supplemental Material, add link to RTLe_V01.xslx.

Dataset License: CC-BY-SA

Keywords: ecotoxicology; pesticide thresholds; benchmarks; environmental risk assessment; effect database; systematic filter application; exceedance rates

1. Summary

In modern agriculture, pesticides are frequently applied to minimize crop losses due to pest organisms, fungal diseases, or the growth of weeds (see Oerke [1], for review). To achieve these agronomic benefits (see Popp et al. [2], for review), pesticides as biologically active substances are regularly and globally released into the environment to treat large areas of croplands [3]. Thereby, pesticides are transported to non-target freshwater systems [4–6]. Hence, it is important to regulate pesticides and to assess the likelihood of unacceptable ecological effects in aquatic ecosystems. For instance, elaborated environmental risk assessments (ERA) are conducted for pesticides in the US, prior to granting permission of use and, at regular intervals thereafter, for reregistration (FIFRA 40 CFR [7], Part 152; [8]). For risk characterization, estimated environmental concentrations are compared to toxicity endpoints from peer-reviewed and non-peer-reviewed laboratory effect studies. Thereby, a pesticide is characterized using different ecologically relevant surrogate species (FIFRA 40 CFR [7], Part

158; [9]). In the US, effect studies are only considered relevant for risk characterizations if they comply to certain validity criteria, as stated in the OCSPP standard test guidelines (e.g., 850.1075, 850.1010, 850.4400, 850.4500 [10–13]) and the US EPA's guidance for the review of open literature studies (e.g., treatments need to be compared to an acceptable control) [14]. The most sensitive, reliable toxicity endpoint is used to calculate risk quotients (RQ), by dividing estimated environmental concentrations by the effect concentrations. The RQ is then compared to a level of concern (LOC, in the acute risk assessment for aquatic animals: 0.5; for plants: 1) [9]. Permission for the use of an active ingredient may only be granted if the RQ does not exceed the respective LOC.

Despite elaborated risk assessment procedures, pesticides were shown to occur frequently in non-target aquatic ecosystems [15,16], where they may have detrimental effects on ecosystem structure [17], functioning [18,19], and biodiversity [20]. Therefore, post-registration monitoring of measured environmental concentrations (MEC) of pesticides in surface waters is importantly needed [16,21,22]. One traditional approach is based on toxicity endpoints of three standard species (algae *Pseudokirchneriella subcapitata*, invertebrate *Daphnia magna*, fish *Pimephales promelas*) that represent different trophic levels (e.g., Malaj et al. [23]). However, these benchmarks consider few test species only and might thus miss the most sensitive, available, and reliable endpoint on which risk evaluations should be based. Especially for some modern insecticides (e.g., neonicotinoids), it was shown that the standard test organism *D. magna* is not an appropriate proxy for the risk of invertebrates as crustaceans are 2–3 orders of magnitude less sensitive than aquatic insects (see Morrissey et al. [24], for review). Even though most likely no toxicity data is available for the most sensitive species occurring in freshwater ecosystems [25], considering all available, reliable toxicity data, such as in US ERA, is a good approach to assess risk.

Regulatory threshold levels (RTL, as introduced in [16]) are, in the present study, directly derived from the effect data (>37 species in total for 143 RTL, see Table S1) provided in regulatory documents. Their use to evaluate the risks of MEC in freshwater systems is also of ecological relevance, as RTL exceedances in the field lead to unacceptable ecological effects [16]. RTL, thus, are a valuable measure to deduce whether aquatic ecosystems are at risk due to pesticide exposure. The compilation of RTL from individual regulatory documents, however, is very time-consuming. For this reason, recent studies using RTL to evaluate MEC have been restricted to a small sample of manually collated RTL (Wolfram et al. [15], n = 32; Stehle and Schulz [16], n = 28). Albeit RTL are found in the regulatory registration documents, they cannot be easily accessed. To identify the RTL for given substances, it is necessary to (i) search online for the most recent regulatory document, (ii) check whether risk assessments were performed and sufficiently reported there, (iii) identify the relevant, reliable endpoint that was used for risk characterization, and (iv) calculate the RTL by applying the respective LOC as a safety factor. Especially steps (i) and (ii) are very time-consuming, as multiple sources where regulatory documents are listed have to be checked [26–29]. Moreover, this extensive manual search inherently lacks the ability to respond to regulatory changes, for instance, in case of a data call-in [FIFRA 40 CFR [7], Part 155] or a pesticide reregistration [FIFRA 40 CFR [7], Part 152]. Thus, it is not possible to adapt RTL collections to changes in the data basis, except via manual labor, which quickly becomes unfeasible due to the plethora of regulated pesticides (i.e., 725 US EPA registration review cases for about 1140 active ingredients published by July 2017 [30]). Given the usefulness of RTL to evaluate environmental risks [15,16], comprehensive RTL collections are urgently needed for large-scale evaluations of MEC. In principal, aquatic life benchmarks (ALB) for pesticides, provided by the US EPA on its web page [28], could be seen as such a collection. However, although ALB are available for a considerable number of pesticides (n = 486), they show two major drawbacks. First, the list of ALB is updated irregularly, second, ALB do partly not relate to the newest regulatory documents. Data errors (i.e., incorrect RTL) have thus been found in about 25% of an evaluated subsample of ALB (n = 143, own evaluation, see Appendix B for details). Therefore, the aim of the present study was to develop and validate a model that derives estimated RTL equivalents (RTLe) for freshwater organisms for a broad number of pesticides and that can be used for scientific risk evaluations. This model is

based on two publicly accessible databases maintained by the US EPA: The US EPA ECOTOXicology knowledgebase (ECOTOX, https://cfpub.epa.gov/ecotox/index.cfm) [31] and the Office of Pesticide Programs Pesticide Ecotoxicity Database (OPP, https://ecotox.ipmcenters.org) [32]. Both databases contain a large number of effect studies (ECOTOX: ~940,000 entries; OPP: ~32,000 entries) and are relevant for the US ERA (also see Section 3.1). However, the quality of the studies listed in the databases varies substantially, so that not all studies reliably describe the toxicity of pesticides. To exclude studies that are unlikely to be used in ERA due to low reliability, filter criteria were formulated based on the provided information on test characteristics (see Tables A1 and A2) and in accordance with official validity criteria [10–14,33]. One rather theoretical approach to systematically filter large aquatic toxicity databases was, for instance, published by Beasley et al. (i.e., SIFT method) [34], and was recently applied to a large set of collected aquatic toxicity data with a focus on industrial chemical risk assessment [35]. In the present study, we developed a similar approach focusing on pesticide risk assessment by translating official validity criteria from standardized test protocols [10–13,33] and the US EPA guideline for the review of open literature studies [14] into an SQL code. Criteria considered from the test protocols included, e.g., study durations, measured effect type, and endpoint type for the respective organism group. From the EPA guideline, e.g., proper control, chemical analysis, and physical units have been extracted. The SQL code was used to query local versions of the ECOTOX and OPP databases. After filter application, the remaining studies were used for threshold estimations (i.e., RTLe) based on the most sensitive, available endpoint, as done during manual RTL derivation [16]. To set up the model, different combinations of filter criteria were tested and evaluated in terms of the precision of estimates (basic, mid, and full model, Tables A1 and A2, Figures A1–A3). To evaluate the precision of the modeled RTLe, the estimated threshold levels were compared to 94 manually retrieved RTL (i.e., the calibration dataset). The model was validated by comparing the RTLe/RTL ratios of the calibration data (n = 94, see Table S1) to the RTLe/RTL ratios of an independent set of pesticides (i.e., validation dataset, n = 49, see Table S1). Furthermore, we demonstrate in a case study using 27 insecticide compounds, how RTLe can be used in actual risk evaluations. The RTLe model resulted in 1283 RTLe, 676 RTLe of pesticides (i.e., fungicides, herbicides, insecticides) plus 607 of other chemicals (see 2.1, Table S2) that might also be of interest for environmental evaluations. This list will be regularly updated to reflect potential changes in the underlying effect databases.

With the present study, we have developed, validated, and applied a dynamic model that estimates a large number of RTLe for freshwater systems based on reproducible validity criteria. With the compiled list of RTLe, comprehensive evaluations of pesticide monitoring data and associated risks in surface waters are enabled for the first time for a large number of pesticides, as well as further chemicals with potential environmental impacts.

2. Data Description

2.1. The Dataset

The dataset contains RTLe for freshwater organisms for 1283 chemicals in ug/L (Table S2). A classification into use types applying the MAGIC graph [36] revealed that 676 of these substances are used as pesticides (herbicides, fungicides, or insecticides). Further, 260 substances were classified as other chemicals (e.g., microbiocides, solvents, plant growth regulators, repellents) and 347 substances remained unclassified. In addition, the CAS number, RTLe, source database, and use type classification are provided (Table 1). In addition to the publication of a Microsoft Excel® spreadsheet containing the compiled data, by which the most recent version of the dataset becomes accessible for the public (https://static.magic.eco/rtle2), all RTLe will be added to the MAGIC graph [36] and are available through the chemical search interface (https://static.magic.eco/rtle1), and updated there regularly.

Table 1. Overview over the published dataset.

Column	Explanation
CAS Number	Chemical abstract service identifier
RTLe	Estimated regulatory threshold level [ug/L]
Database	Source of the presented RTLe, i.e., ECOTOX database, OPP database
Classification	Fung = fungicides, Herb = herbicides, Ins = insecticides, Other = other category according to MAGIC graph [36] classification, Unclassified = no classification

2.2. Data Quality

The RTLe from the calibration dataset (n = 94, Table S1) were validated against an independent sample of RTL (n = 49, Table S1) to test the model's performance. Model transferability was confirmed by non-parametric hypothesis testing (Wilcoxon rank sum test: ECOTOX database: W = 2148.5, p = 0.628; OPP database: W = 1527.5, p = 0.474), which showed no significant difference in predictive accuracy between calibration and validation data (Figure 1, Figure A5). The proportion of RTLe that correctly estimated the RTL after combining the queries of both databases (Figure 1) was 48.9%. Further, 95.6% of RTLe were within ± 1 orders of magnitude, while 87.6% lay within ± 0.5 orders of magnitude (i.e., a factor of 3.2). Thus, for the vast majority of data, minimal deviance from official RTL can be assumed, underlining the robustness of the presented model. Estimates for only few substances (n = 5, i.e., tebufenozide, fenthion, heptachlor, methyl-parathion, sulfometuron-methyl) underestimated the RTL by more than 0.5 orders of magnitude. Whenever the RTL was underestimated, this means that a more sensitive endpoint was listed in the databases for which the applied exclusion criteria did not apply. This might be due to expert judgements, as these remaining endpoints after filter application were probably omitted during regulatory risk characterizations by risk assessors for reasons currently not reflected by the model (as not all relevant test characteristics are consistently encoded in the databases). As this was the case for only a few substances (Figure 1, RTLe/RTL < 1), the overall probability to underestimate the RTL remains low.

Albeit model estimates were on average significantly higher (based on 95% CIs) than the original RTL (bootstrapped mean with 95% CI: 1.21 (1.04–1.42 95% CI; median = 1)), RTLe would rather underestimate the risk if used in risk evaluations, resulting in a low likelihood of false-positive risk indications. Since only 88 (ECOTOX database) and 78 (OPP database) of the relevant, most sensitive endpoints for the 143 manually retrieved RTL were included in the respective databases (i.e., exact match between the endpoints used for RTL and RTLe compilations), correct estimates were inherently not achievable for all pesticides. However, for some pesticides, deviations were only marginal (deviation < ± 0.1 orders of magnitude, excluding 0, n = 16 (ECOTOX database) and n = 26 (OPP database). In some cases, the underlying endpoints might still originate from the same study (identifiable by the reported study ID and same test organism), e.g., in case of tribufos, methiocarb, and formetanate HCl. For instance, the minor deviations could result from the rounding of endpoints if different numbers of digits behind the comma were reported in the document and the database (as likely for formetanate HCl; 0.09 parts per million (ppm) in the document [37], and 0.0868 parts per million in the OPP database). Larger deviations could, however, be based on recalculations of endpoints by EPA staff (e.g., to account for the percentage of active ingredient if original studies were conducted with formulations). Additionally, it became apparent that endpoints were more likely to not be included in the respective database the newer the regulatory document was (see Figures A6 and A7). Hence, there appears to be a time-lag between the update of the databases and the publication of regulatory documents.
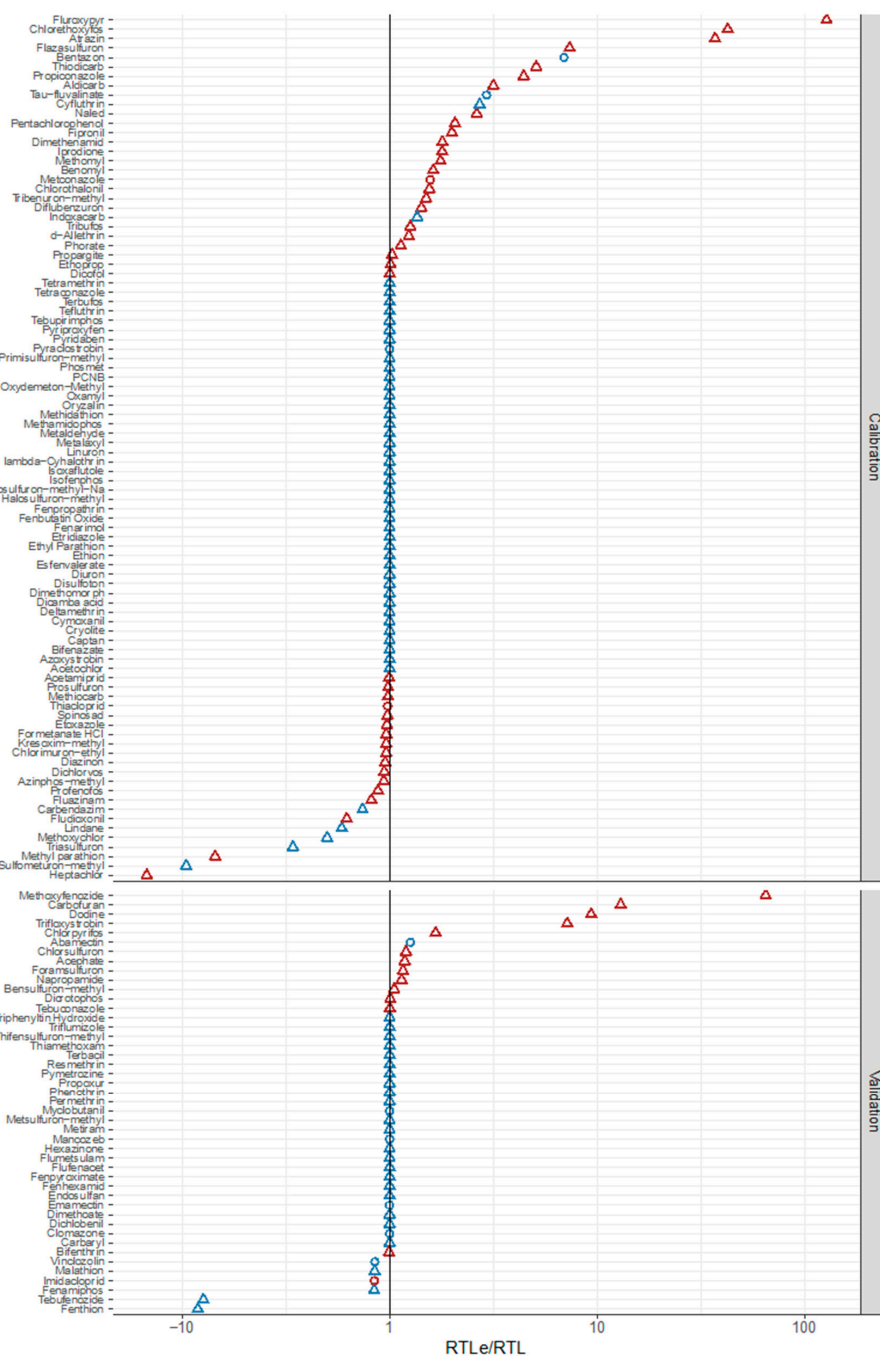
Figure 1. Precision of the final regulatory threshold level equivalent (RTLe) model expressed as RTLe to RTL ratio (log10-transformed) for the calibration and validation datasets (n = 137). The one line indicates where RTL are estimated correctly (RTL = RTLe). A ratio < 1, shows an underestimation of the RTL by the RTLe, a ratio > 1 an overestimation. Triangles indicate whether RTLe were based on data provided in the Office of Pesticide Programs Pesticide Ecotoxicity Database (OPP); circles refer to the US EPA ECOTOXicology knowledgebase (ECOTOX). Not all relevant endpoints that were listed in the regulatory documents were also listed in the respective databases. Pesticides, with included endpoints are displayed in blue, pesticides without the listed, most sensitive, relevant endpoint in the respective database are displayed in red.

If only pesticides with a relevant, most sensitive endpoint that is also included in the database were considered, the precision of estimates improved significantly. A strong association between the two variables was indicated as the bootstrapped mean ratio approached 1 (mean ratio RTLe/RTL

of 0.94 (0.84–1.05 95% CI; median = 1)) and minimal deviations for the majority of pesticides were observed (Figure 1). Additionally, RTLe for pesticides whose relevant, most sensitive endpoint for RTL compilation was not included in the databases tend to overestimate the RTL (i.e., RTLe/RTL > 1) with a bootstrapped mean ratio with 95 CI of 1.75 (1.27–2.48 95% CI; median = 1.16). These additional analyses demonstrate a good performance of the model particularly for the pesticides with the included relevant endpoint for RTL compilation, indicated by a high precision of estimates. For relevant, most sensitive endpoints from the regulatory document, that were used for RTL compilation but were not included in the respective databases, even though it was inherently not possible to obtain correct estimates, RTLe rather overestimated the RTL. Hence, the use of these RTLe would not result in false positive risk estimations. Thus, overall, the model precision is high with a tendency to slightly overestimate the RTL. Consequently, based on the uncertainty associated with these modeled thresholds, ecological effects in surface waters could also occur at concentrations below the respective RTLe, and are very likely if RTLe are exceeded (cf. Stehle and Schulz [16]).

2.3. Case Study

The model was applied for 27 insecticides from an extensive dataset of quantified (i.e., ≥ limit of quantification) insecticide surface water concentrations (i.e., MEC, n = 3001, 1962–2017) in the United States described in detail in Wolfram et al. [15]. The RTLe for the insecticides used in this case study are, in terms of model precision, a representative sample of all RTLe, for which the model was calibrated and validated (Wilcoxon rank sum test: W = 1429, p = 0.749). Risk evaluation (i.e., the calculation of exceedance rates) of MEC using RTLe showed only marginal differences of 3.8% (95% CI 2.2–5.4) if compared to the evaluation of MEC using RTL. In detail, bootstrapped median exceedance rates lay at 55.8% (95% CI 51.8–59.8) and 52.0% (95% CI 48.0–56.0) if insecticide MEC were compared to RTL and RTLe, respectively (Figure 2). No statistically significant difference (based on 95% CIs) between the exceedance rates were observed if MEC were compared to RTL or RTLe. This underlines the model's ability to accurately identify environmental risks. However, using RTLe, the risk was slightly underestimated, such that using the modeled thresholds resulted in a less conservative risk evaluation (blue line, Figure 2). In conclusion, we found it being representative of the entirety of RTLe and providing comparable risk evaluation based on actual field measurements.
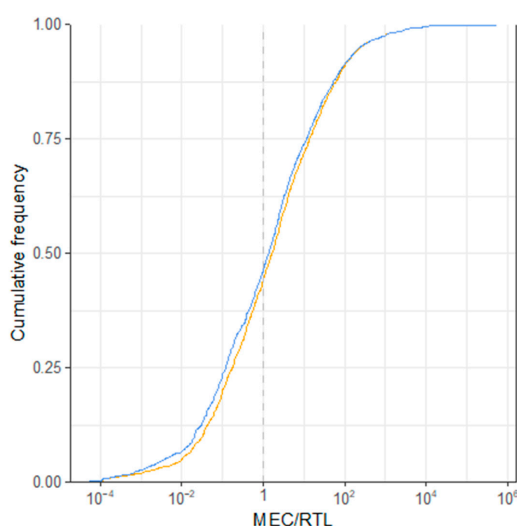

Figure 2. Cumulative risk distribution for measured environmental concentrations (MEC, n = 3001) of 27 insecticides in US surface waters (see Wolfram et al., 2018) compared to RTL (orange line) and RTLe (blue line), respectively. The vertical dashed line indicates where MEC equals RTL or RTLe. Above a ratio MEC to RTL or RTLe of 1, MEC exceeded the threshold. The proportion of exceeded MEC (log10 (MEC/RTL) > 1) lies at 0.56 (bootstrapped 95% CI 0.52–0.60) and 0.52 (bootstrapped 95% CI 0.48–0.56) for RTL and RTLe, respectively.

3. Methods

3.1. The Databases

The present study was based on two publicly accessible effect databases maintained by the US EPA (i.e., ECOTOX database and OPP database, see Figure A1), as introduced in Section 1. The ECOTOX database contains predominantly peer-reviewed, open literature studies of ecotoxicological effects on a variety of test organisms for individual substances. The OPP database contains data from peer-reviewed and non-peer-reviewed studies that were previously reviewed by the US EPA for risk characterizations within the process of pesticide regulations, also for a variety of test organisms. The OPP database is more aggregated than the ECOTOX database (i.e., contains less information for individual studies). Therefore, the possibility to formulate filter criteria (Tables A1 and A2) to select only valid endpoints was limited to the provided information. However, since EPA staff already rated the reliability of these studies as invalid, supplemental and core, relevant endpoints can be selected by excluding invalid studies, as only supplemental and core studies are used for ERA by the US EPA [9]. Both databases were downloaded and imported into a relational database management system (PostgreSQL, pgAdmin 4.1; latest database updates: ECOTOX: 06/2019; OPP: 03/2017). Data was analyzed using R [38], and the following packages [39,40].

The test medium was used as a criterion to distinguish freshwater organisms from saltwater/estuarine organisms to enable the estimation of RTLe for freshwater systems. As the OPP database does not provide information on the test medium, unlike the ECOTOX database (e.g., freshwater, saltwater), the typical, predominantly used test media were retrieved from the ECOTOX database by joining species and media tables by species name. All test media for species that were not listed in the ECOTOX database but occurred in the OPP database (n = 61), were inferred from the predominant habitat specified, as specified in additional sources (e.g., WoRMs database (http://www.marinespecies.org), Algaebase database (https://www.algaebase.org), Stephen et al. [41–43]).

3.2. Collection of RTL

A list of 143 RTL (updated by October 2018) used for model validation and calibration was compiled by reviewing the respective, most recent regulatory documents provided by the US EPA from different sources [26–29]. Active substances were either searched by CAS number or name to unambiguously identify each test substance. Pesticides for the present studies were chosen based on previous work [15,16], use data [44] and the availability of regulatory documents on the searched sources covering different important use types (i.e., fungicides, herbicides, and insecticides). Once the most recent documents were identified and downloaded, RTL were computed similar to the methods described in Stehle and Schulz [16] as follows: First, the valid ecological effect data for freshwater organisms, that was used for risk quotient (RQ) calculations within the respective document, was identified. As the model was supposed to be calibrated for active ingredients, only studies performed with the active ingredient of a pesticide were selected. Further, only definite endpoints (without the qualifiers ">" and "<") were considered to be relevant. If no information on RQ calculations were provided but sufficient ecological effect data was listed, the relevant endpoint was chosen based on the US EPA's classification of endpoints as supplemental and core (excluding invalid studies). This follows the procedure of endpoint selection by the US EPA during the screening of available data for ERA [9]. As a final verification of the manually selected endpoint, written justifications of endpoint selections, if available, were checked to confirm the identified species and effect endpoint. Finally, as the most sensitive species drives regulatory decisions [9], the lowest, valid endpoint was selected for RTL computation. These RTL were calculated by multiplying the selected, most sensitive, relevant endpoint by the respective level of concern (LOC; 0.5 for acute aquatic animal studies, 1 for aquatic plant studies) [9]. The final dataset of RTL was divided randomly into a calibration dataset (n = 94) and a validation dataset (n = 49) to set up the model.

3.3. Model Building and Calibration

To set up the model, a basic, mid, and full model were tested differing in the strictness of applied filter criteria for each database separately. The basic model was set up containing elementary filter criteria to define a basic dataset (basic model, see Tables A1 and A2, and Figure 3). Then, based on the most sensitive, remaining endpoint per substance, RTLe were calculated for the substances of the calibration dataset (see Section 1. Summary). The application of the basic models would result in a large underestimation of the RTL (i.e., overestimating risks), as for 90% and 66% of the substances for the ECOTOX and OPP database, respectively, the RTLe would be estimated too low if compared to the RTL. At the same time, the proportions of correct estimates (2% and 21%) and overestimates (8% and 13%) were rather low, such that this query would not result in accurate RTL estimates. To improve model precision, additional filter criteria (Tables A1 and A2) were formulated based on quality standards [10–14,33], as introduced in Section 1, to identify and exclude data with a low reliability (i.e., which are likely to be rejected during regulatory risk assessments). A mid model containing essential criteria (mid model, n = 4, for the ECOTOX and OPP database, respectively; see Tables A1 and A2) was set up and RTLe were calculated and compared to RTL to evaluate the precision of estimates. Applying the mid model, model precision improved, as 28% and 42% of the RTLe were estimated correctly, 54% and 34% of RTLe underestimated their RTL and 17% and 24% of RTLe exceeded their RTL, for the ECOTOX and OPP database, respectively. For further refinements, additional criteria were formulated to build a full model, which contains a total of 17 and 8 consecutive filter criteria for the ECOTOX and the OPP database, respectively (full model, Tables A1 and A2). Here, the model performance could again be improved, as 43% and 47% of the RTLe were estimated correctly, only 25% and 24% of RTLe underestimated their RTL, and 32% and 29% of RTLe exceeded their RTL, for the ECOTOX and OPP database, respectively (see Section 2.2 Data quality and Section 2.3 Case study for further details on the precision of estimates). Thus, calibration aimed at increasing the proportion of correct estimates (RTLe/RTL = 1) while decreasing the proportion of substances for which the RTLe would underestimate the true RTL (RTLe/RTL < 1, see Figure 3). Consequently, the comparison of RTLe to MEC would be unlikely to result in a false positive overestimation of the risk.
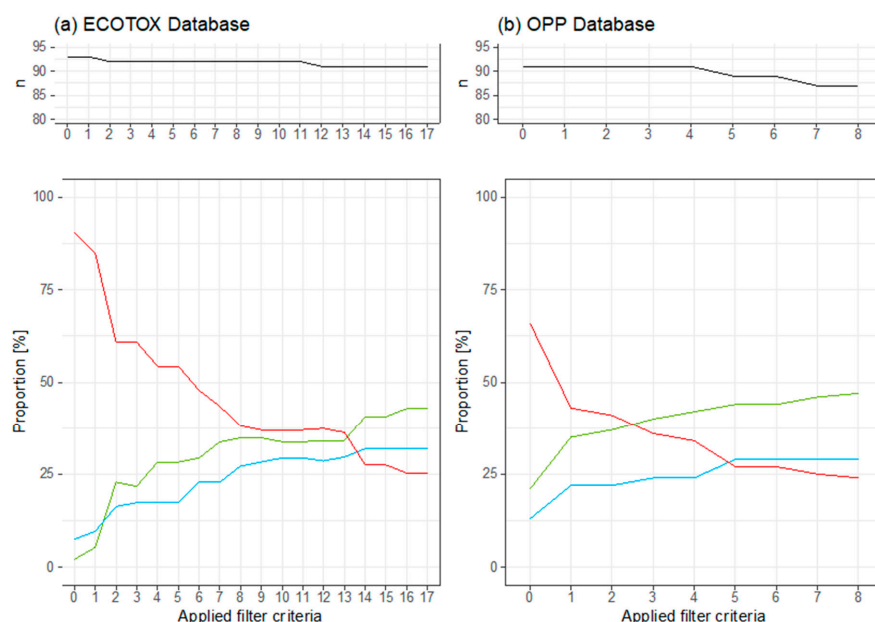

Figure 3. Overview of the influence of applied filter criteria (Tables A1 and A2) on the remaining number of pesticides for which RTLe could be estimated for the ECOTOX database (**a**) and the OPP database (**b**), respectively. Upper panels show the remaining number of pesticides, lower panels the proportions of RTLe that were estimated correctly (green lines), estimated too high (blue line), or too low (red line).

As a consequence of filter refinements, the application of each additional filter step resulted in a minimal loss of endpoints (see Figures A2–A4) of lower reliability. In this process, for some substances, all data was lost, such that no final RTLe could be estimated (calibration dataset: n = 94 pesticides; null model estimates: n = 93 and 91 pesticides; full model estimates: n = 91 and 87 pesticides, respectively, for the ECOTOX and OPP database). This was a trade-off, where improving the model precision was preferred over maximizing the generation of RTLe.

3.4. Model Validation and Join of ECOTOX and OPP Query Results

The model was validated on two levels: (i) The measure of precision (log10 (RTLe/RTL)) for calibration and validation data was compared by hypothesis testing to test whether the precision of estimates was statistically different when the model was applied to an independent set of pesticides. In case of non-normality and heteroscedasticity, non-parametric testing was conducted (i.e., Wilcoxon Rank Sum test, see Section 2.2 for results). Then, (ii) it was tested whether external factors (i.e., number of available endpoints prior to filter application, publication year of the most recent regulatory document, use type of the pesticide) had an influence on the model precision by either hypothesis testing or testing for correlations. The number of available endpoints per substance had no influence on the precision of estimates (Spearman's rank correlation rho: ECOTOX database: S = 450,422, p = 0.389, rho = −0.074; OPP database: S = 308,287, p = 0.558, rho = 0.052). Further, the test of a potential association between the publication year of the regulatory document and the measure of model precision revealed that RTL from newer studies were more likely to be overestimated (significant for ECOTOX database estimates: S = 316,508, p = 0.004, rho = 0.245; not significant for OPP database estimates: S = 300,311, p = 0.391, rho = 0.077). Further, the RTL based on endpoints reported in newer publications were less likely to be included in the respective database (see Figures A6 and A7). Thus, there appears to be a delay in the update of the databases. Finally, the test for differences between the estimates for different use types (i.e., herbicides, fungicides, and insecticides) revealed no significant differences (Kruskal–Wallis rank sum test: Kruskal–Wallis chi-squared: 3.64, df = 2, p = 0.16). This was confirmed by a post-hoc pairwise Wilcoxon rank sum test. Thus, it was concluded, that the precision of the model does neither rely on the number of endpoints available for the respective pesticide, nor on the use type. However, dependencies became apparent for the publication year and the presence of the most sensitive, relevant endpoint used for RTL compilation in the respective database. That means that RTL were more likely to be estimated too high for newer publications and non-included endpoints (see Figures A6 and A7).

After the model was successfully validated, the final dataset of RTLe (Table S2) was compiled by pooling the calibration and validation data and merging the estimates from the different databases: The ECOTOX data was joined to the OPP data, whereby all RTLe from the OPP database were kept. Here, OPP estimates were preferred over ECOTOX estimates, as the classification into core, supplemental, and invalid studies allowed to exclude invalid studies with a higher degree of confidence than in the ECOTOX database. Thus, the OPP dataset was complemented with ECOTOX data, for which no OPP-RTLe was available.

3.5. Bootstrapped Model Precision

Model precision was estimated by bootstrapping mean ratios (RTLe/RTL, n = 1000), where a mean close to 1 would indicate a high model precision. Vice versa, the larger the departure from 1, the higher the imprecision. While a mean ratio < 1 indicates that estimates would rather be too low and the model would result in too sensitive thresholds, a mean ratio > 1 would indicate conservative threshold estimates. Ratios were sampled from the final dataset (combination of OPP and ECOTOX estimates) with replacements of the same size as the original dataset and with 1000 repetitions (Monte Carlo resampling) following methodologies as outlined in [45,46]. Then, from the bootstrapped samples of ratios, the mean ratio, median ratio, and 95% confidence limits as 0.025 and 0.975 centiles (percentile method, e.g., [45]) were computed.

Further, it was tested whether the endpoints leading to the RTL for all substances were included in the respective databases such that correct estimates would be theoretically possible. By again bootstrapping mean ratios of RTLe to RTL, it was tested whether model precision improved if only pesticides were considered, whose relevant endpoint actually occurred in the respective database.

3.6. Model Application

For the final model queries, the seventeenth and eighth filter criteria for the ECOTOX and OPP databases, respectively, were removed as it was no longer necessary to restrict the results to the years of publication. These filter criteria were only used for calibration and validation of the model, as potentially occurring endpoints introduced into the databases after the publication of the regulatory documents could not yet be included to regulatory risk characterizations and would thus bias the model precision.

The final list of RTLe will be updated regularly and included to the magic graph (see https://static.magic.eco/rtle1 and https://static.magic.eco/rtle2), as the ECOTOX database is updated four times a year. Simultaneously, the OPP is checked for updates, whenever the ECOTOX database is updated such that the dataset, on which the model and thereby the RTLe are based, cover the latest studies added to the effect characterizations.

Here, the risk assessment procedure was applied to all substances listed in the ECOTOX database and OPP database, as both databases are relevant to the risk assessment process of pesticides. However, beside the 676 RTLe for pesticides, additional RTLe for chemicals of other use types are returned by the model. These might be of special interest to complement scientific risk evaluations, as the model enables the direct comparison of pesticides to other use types based on uniformly derived thresholds.

3.7. Case Study

To evaluate the likelihood to over- or underestimate the risk if exceedance rates are calculated by comparing MEC to RTLe instead of RTL, we performed a case study based on MEC ≥ quantification limits as reported by Wolfram et al. [15] (n = 3001 MEC for 27 insecticides). First, risk distributions [16] obtained by using RTL and RTLe, as well as the resulting exceedance frequencies, were compared descriptively. In addition, to test for significant differences in median exceedance rates, we used bootstrap methods to calculate medians with 95% confidence limits [46,47]. Therefore, we sampled 500 MEC with replacements and 1000 repetitions from this dataset to estimate median exceedance rates. Then, for each resampled dataset, the proportion of MEC exceedances was calculated for RTL and RTLe, respectively. From the resulting bootstrapped distributions of exceedance rates, the median with confidence limits were calculated by taking the 0.025, 0.5, 0.975 centiles (i.e., percentile method, e.g., [45]). Additionally, the median difference (in percent points) between the exceedance rate of MEC compared to RTL and RTLe, respectively, was bootstrapped with confidence limits.

To ensure that RTLe used in this case study (n = 27) are representative for the entirety of RTLe from the pooled data (i.e., calibration and validation data, n = 137) in the study, we used hypothesis testing (i.e., Wilcoxon Rank sum test) to test for significant differences between the aforementioned data subset and pooled dataset.

4. User Notes

RTLe can be used to evaluate pesticide MEC and to calculate exceedance rates (for methods see [15,16]). They are intended for use in large scale analyses over a large number of substances where, despite deviations for individual substances, a reasonable statistical certainty can be achieved (see Section 2.2 Data quality). For studies with only a few substances, consultation of the respective regulatory documents should be considered.

Supplementary Materials: The following are available online at http://www.mdpi.com/2306-5729/4/4/150/s1. Table S1: aux_rtl.xslx; Table S2: RTLe_V01.xslx.

Appendix A

Table A1. ECOTOX database filter criteria.

Filter Criteria	Description	Model
0	Basic filter criteria, that select studies with a concentration larger than 0, convertible to ug/L, and conducted with freshwater medium. Further, a concentration must be specified, and not include "NR", "ca", ">", "~", or "x".	Basic
1	The test type needs to be specified as "acute" or "not coded".	Mid
2	Only acute endpoint types are selected, that describe a concentration at which an effect occurs for 50% of the tested organisms (i.e., LC_{50}, EC_{50}, IC_{50})	Mid
3	For animals, mortality and intoxication are allowed as effects; for plants, growth, reproduction, and population effects are selected.	Mid
4	The exposure duration should be one of the following: For animals: 48, 72, 96 h, 2, 3, 4 d for plants: 120 h, 5, 7, 14 d	Mid
5	Exclude studies with multiple measurements.	Full
6	Exclude studies with effect comments.	Full
7	The test substance purity should have a minimum of 70% active ingredient or specified as not coded not reported.	Full
8	Tests should be conducted with active ingredients, such that the concentration type should be the active ingredient or specified as not coded or not reported.	Full
9	Only laboratory studies are considered relevant.	Full
10	Certain formulations are excluded by name (e.g., IN-MP819, INJT333, DPX-KN125, INU8E24, INUYG24)	Full
11	The organism's source should not be specified as wild or multiple.	Full
12	Only definitive endpoints are considered relevant.	Full
13	The test method must be specified.	Full
14	A chemical analysis must be conducted.	Full
15	The test should have been conducted with a proper control.	Full
16	Test organisms characterized with '> 24 h', '>24 h', or 'Fry' and tests conducted with the larval life stage of *Macrobrachium rosenbergii* are excluded.	Full
17	The publication year of the regulatory document needs to be older or equal to the reference year for the respective study. [1]	Full

[1] Only applicable for calibration and validation of the data.

Table A2. OPP database filter criteria.

Filter Criteria	Description	Model
0	Basic filter criteria, that select studies with a concentration larger than 0, convertible to ug/L, and conducted with freshwater medium. Further, a concentration must be specified, and not include "NR", "ca", ">", "~", or "x".	Basic
1	Only acute endpoint types are selected, that describe a concentration at which an effect occurs for 50% of the tested organisms (i.e., LC_{50}, EC_{50}, IC_{50})	Mid
2	The exposure duration should be one of the following: For animals: 48, 72, 96 h for plants: Lower or equal to 336 h	Mid
3	Studies should be rated as supplemental or core.	Mid
4	Only definitive endpoints are considered relevant.	Mid
5	Certain formulations, or metabolites are excluded by name (e.g., IN-MP819, INJT333, DPX-KN125, weed killer, product, formulation, metabolite)	Full
6	The guideline should not be specified as NG (no guideline)	Full
7	The test substance purity should have a minimum of 70% active ingredient.	Full
8	The publication year of the regulatory document needs to be older or equal to the reference year for the respective study. [1]	Full

[1] Only applicable for calibration and validation of the data.
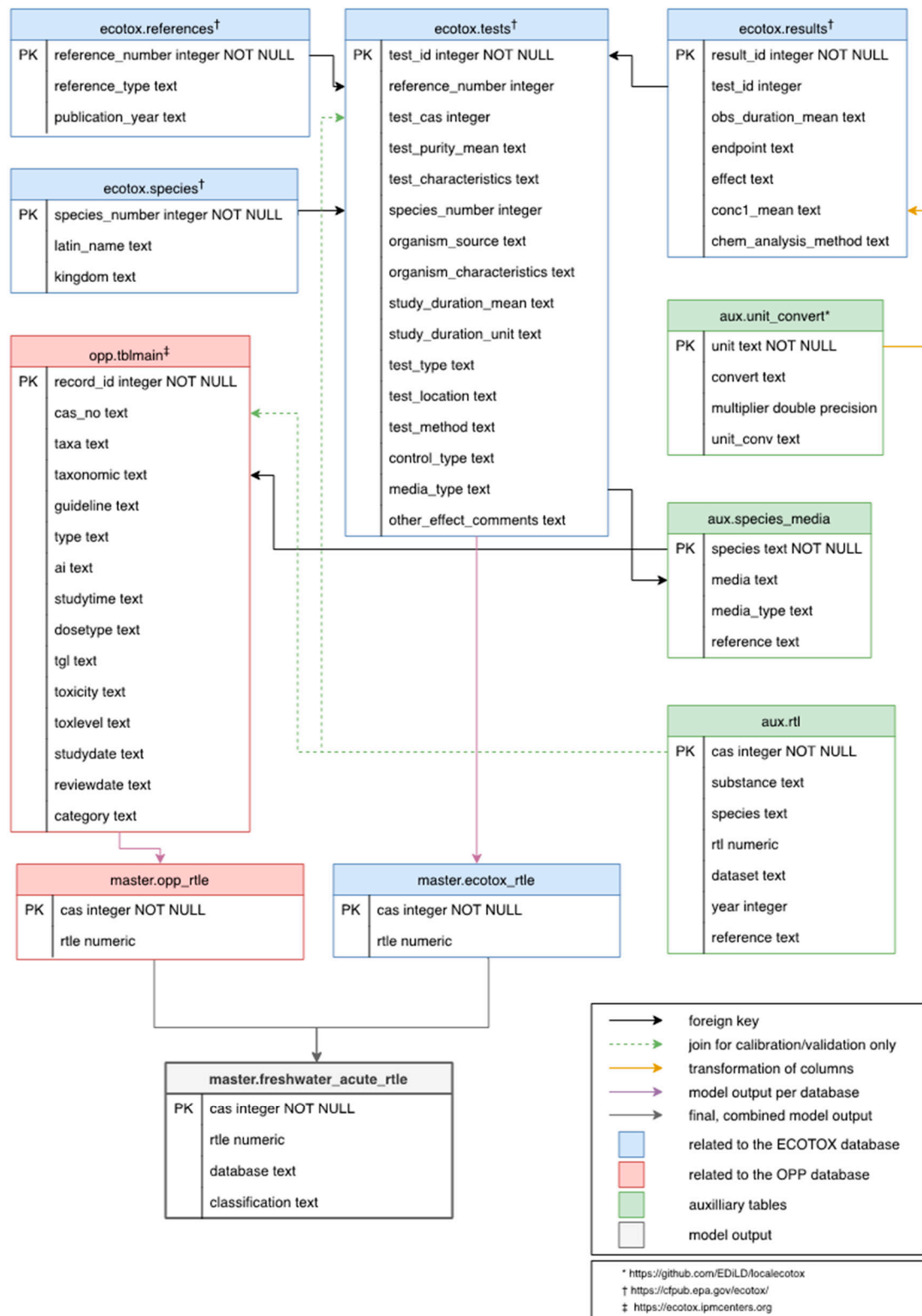
Figure A1. Overview on the relationships between the databases and tables on which the model is based. The data from the master.freshwater_acute_rtle table (S2, file name: RTLe_V01.xslx) and the aux.rtl table (S1, file name: aux_rtl.xslx) are published as Supplemental Material.
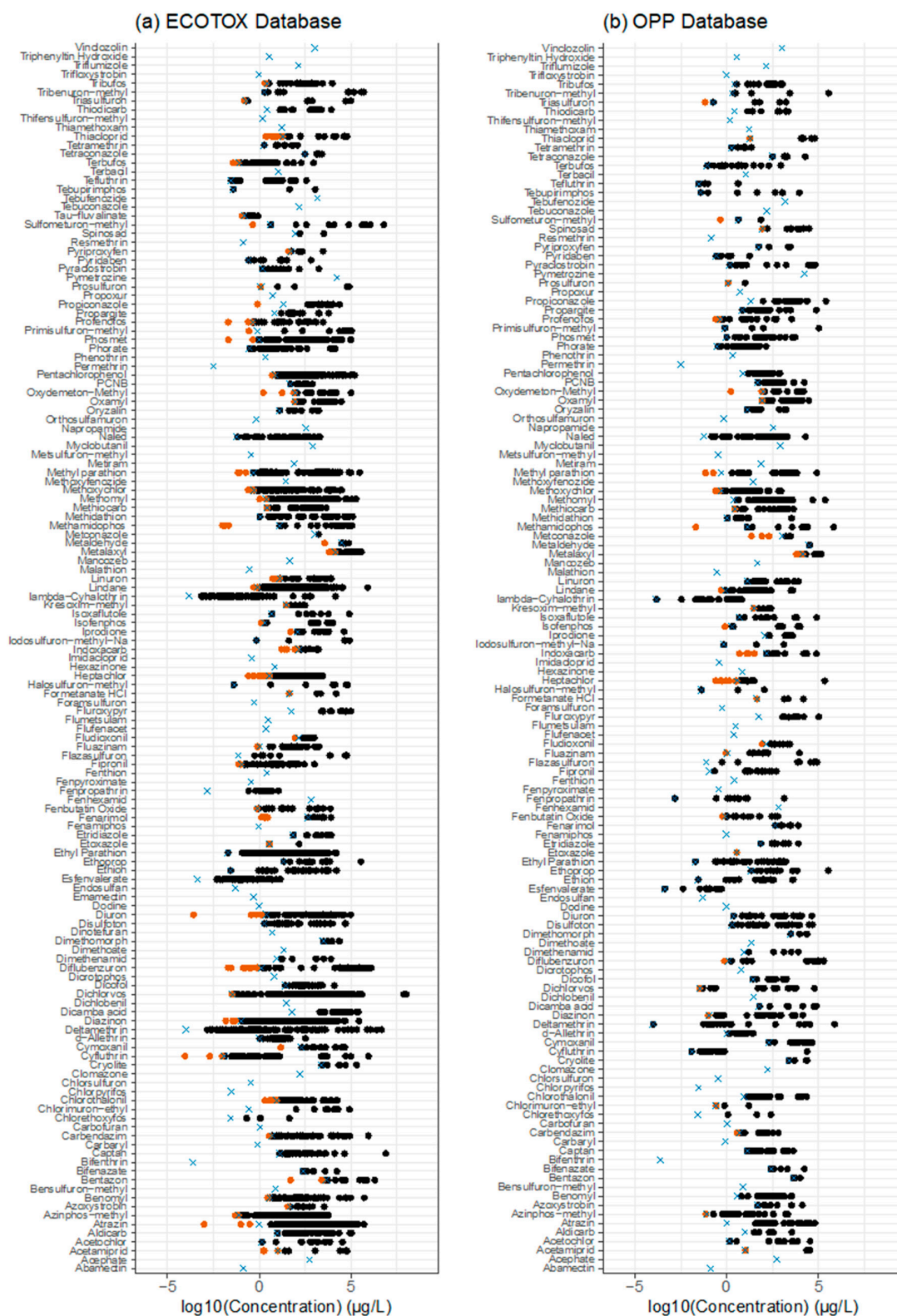
Figure A2. Distribution of remaining endpoints after basic model application for the ECOTOX database (**a**) and the OPP database (**b**). Orange dots indicate endpoints that are more sensitive than the relevant endpoint used for RTL compilation. Blue crosses indicate the relevant endpoint for RTL compilations and black dots indicate endpoints that are less sensitive than the relevant endpoint used for RTL compilations.
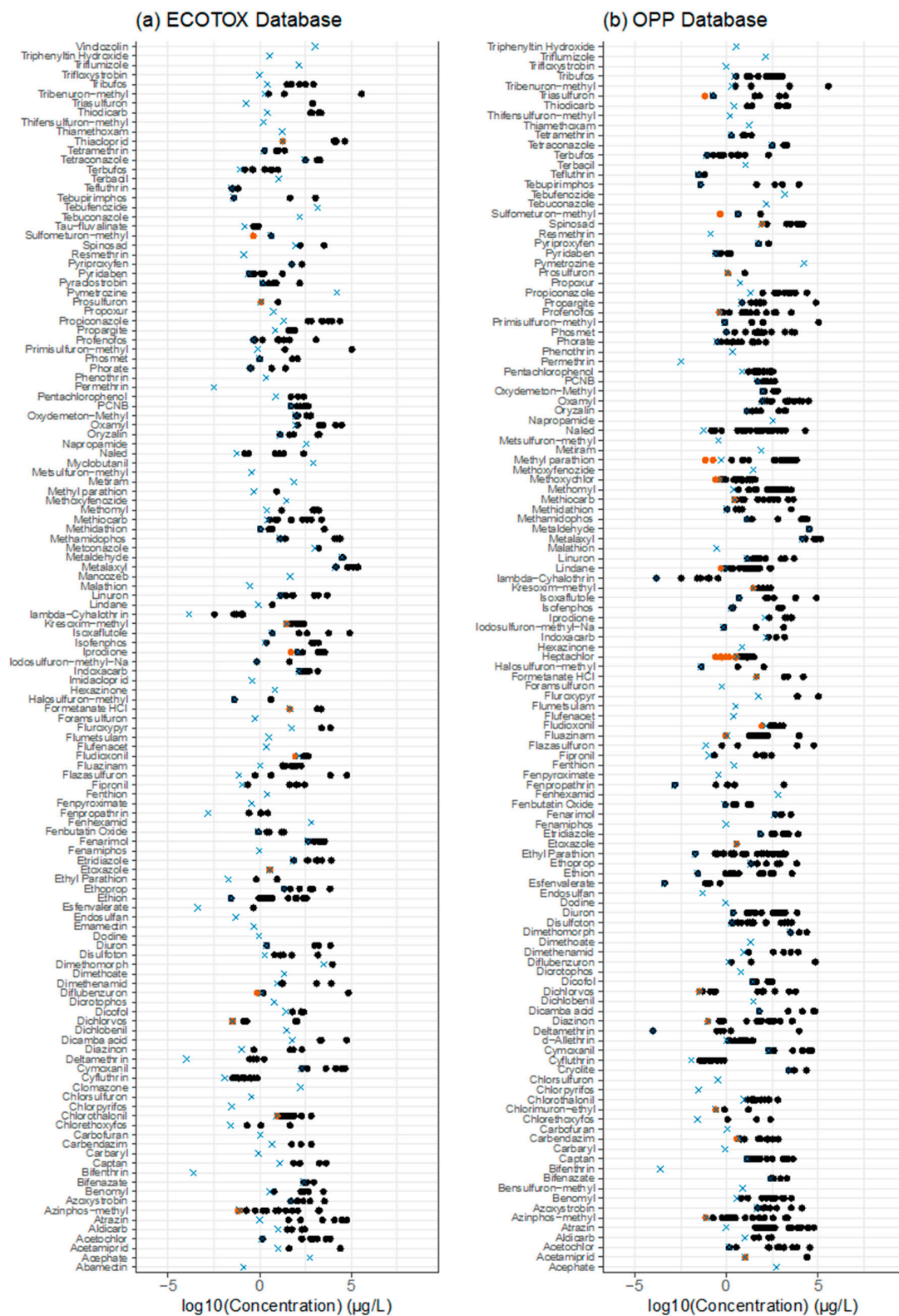
Figure A3. Distribution of remaining endpoints after mid model application for the ECOTOX database (**a**) and the OPP database (**b**). Orange dots indicate endpoints that are more sensitive than the relevant endpoint used for RTL compilation. Blue crosses indicate the relevant endpoint for RTL compilations and black dots indicate endpoints that are less sensitive than the relevant endpoint used for RTL compilations.

Figure A4. Distribution of remaining endpoints after full model application for the ECOTOX database (**a**) and the OPP database (**b**). Orange dots indicate endpoints that are more sensitive than the relevant endpoint used for RTL compilation. Blue crosses indicate the relevant endpoint for RTL compilations and black dots indicate endpoints that are less sensitive than the relevant endpoint used for RTL compilations.
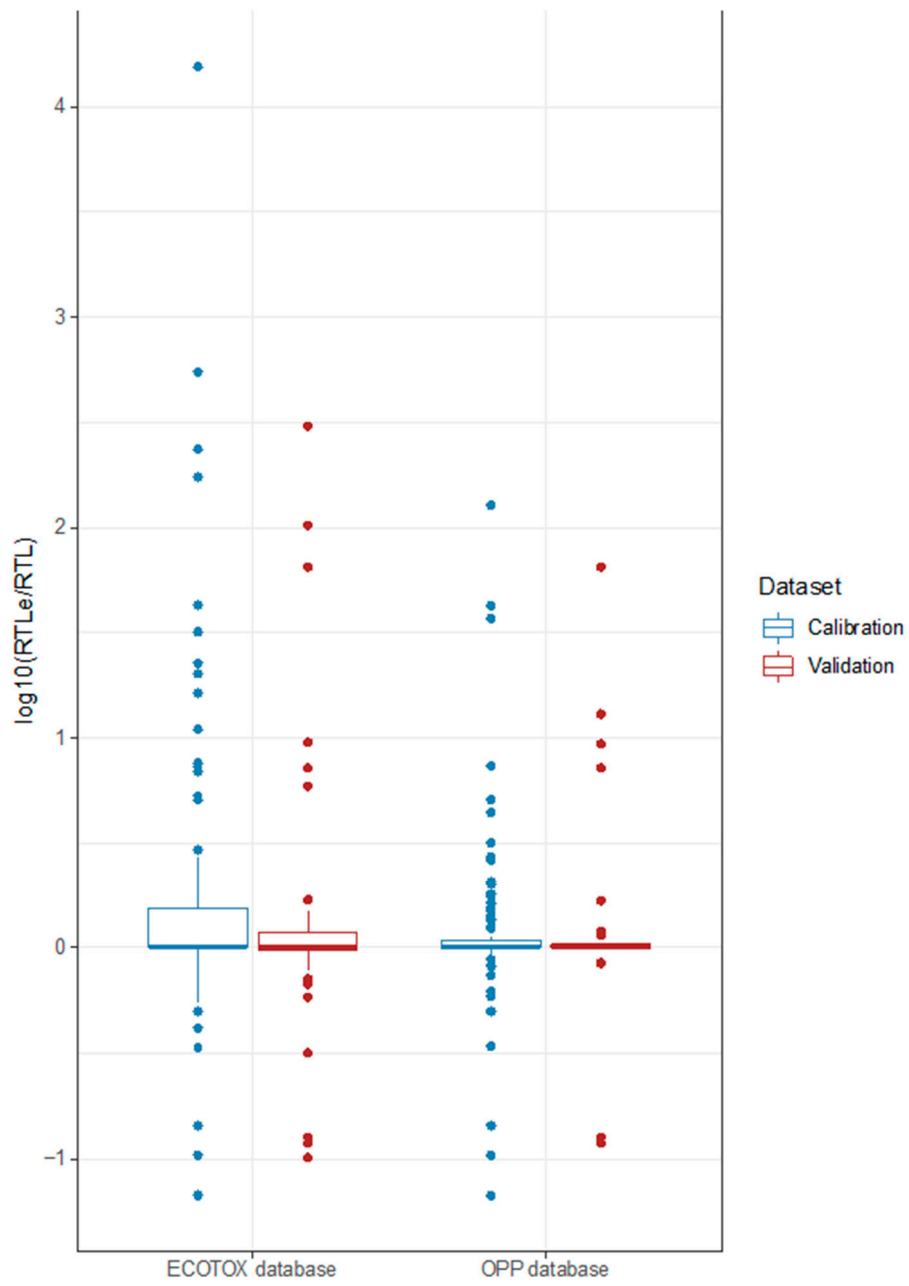
Figure A5. Comparison of model precision (expressed as log10 ratio of RTLe to RTL) for calibration and validation data based on ECOTOX and OPP database estimates.
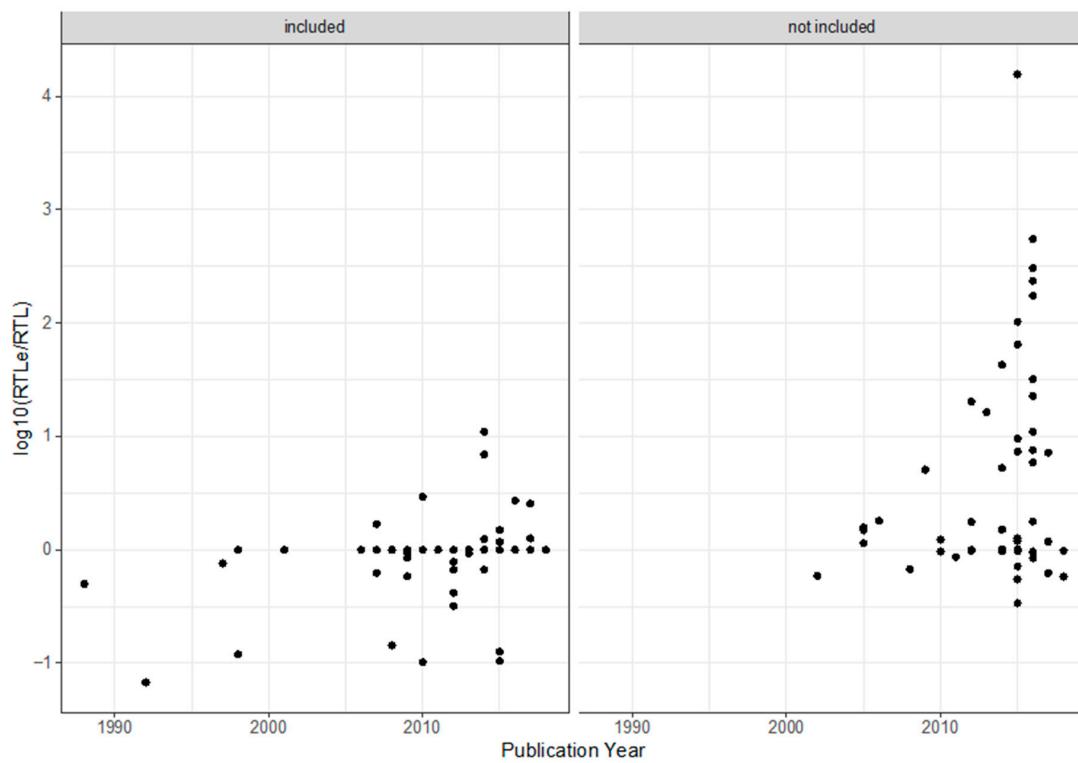
Figure A6. Comparison of model precision (expressed as log10 ratio of RTLe to RTL) for ECOTOX database estimates in dependence on the publication year of the regulatory document for RTL that were included in the database and not included.
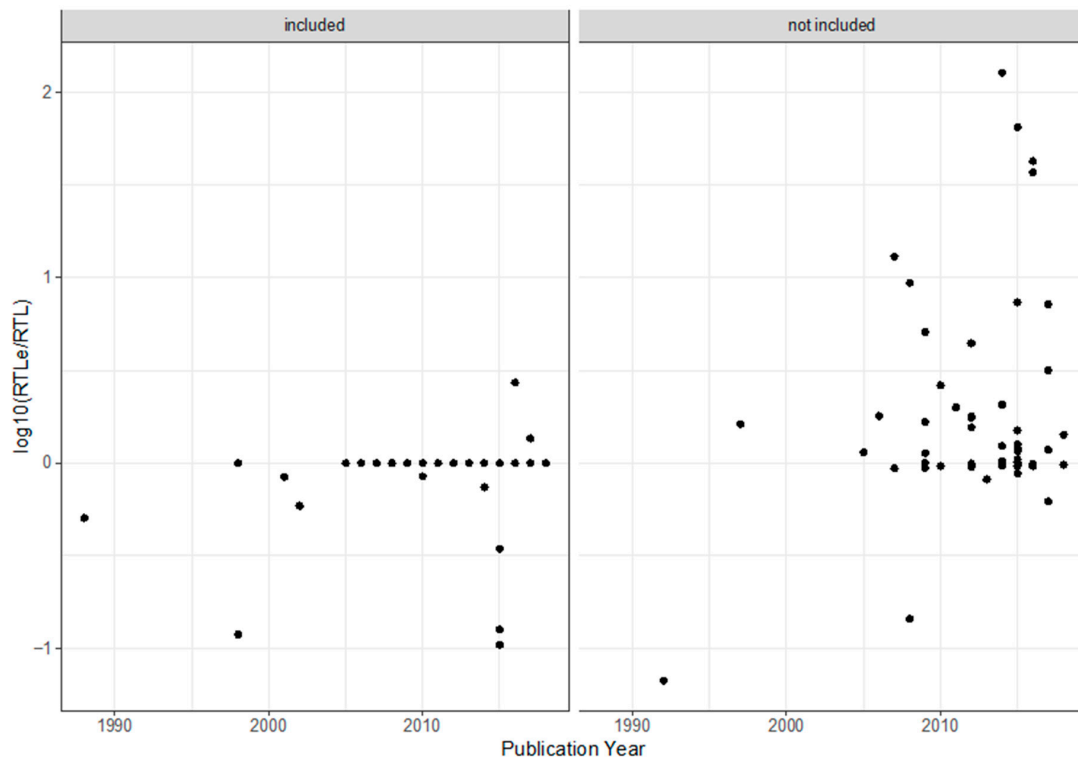

Figure A7. Comparison of model precision (expressed as log10 ratio of RTLe to RTL) for ECOTOX database estimates in dependence on the publication year of the regulatory document for RTL that were included in the database and not included.

Appendix B. Evaluation of Aquatic Life Benchmarks (ALB)

In autumn 2018, it was tested whether ALB could be used for comprehensive risk evaluations instead of RTL, as they are provided by the US EPA, theoretically originate from the same sources as RTL (i.e., regulatory documents) and are also compiled by comparing effect endpoints with the respective level of concern [28]. Therefore, we tested whether the list of retrieved RTL for 143 pesticides agrees with the provided ALB for the same pesticides. It became apparent that for the 143 pesticides, approximately one third of ALB were either wrong or not included in this list (in October 2018). A recent check (October 2019) of the ALB revealed that the US EPA has updated the list of ALB. Thus, it was re-assessed whether the incorrect or outdated ALB were corrected, which was the case for 10 of the 143 reviewed pesticides. Still, after the most recent update, for a total of 25% of the evaluated pesticides, the ALB either deviate from the manually retrieved RTL from the most recent regulatory documents or no ALB is reported for the evaluated pesticides (n = 26 and 10 pesticides, respectively). Thus, the list of ALB, even though becoming more comprehensive, largely relies on regular updates. Additionally, a time lag between the publication and the most recent regulatory document and the referenced document in the ALB data collection became apparent (for wrong ALB, median time lag = 4 years, mean = 3.8 years). Finally, ALB were available for 486 unique pesticides (identified by CAS number), representing only a subsample of the actually reviewed active ingredients by the US EPA (i.e., 1140 published by July 2017 [30]).

References

1. Oerke, E.C. Crop losses to pests. *J. Agric. Sci.* **2006**, *144*, 31–43. [CrossRef]
2. Popp, J.; Pető, K.; Nagy, J. Pesticide productivity and food security. A review. *Agron. Sustain. Dev.* **2013**, *33*, 243–255. [CrossRef]
3. Food and Agriculture Organization of the United Nations (FAO). FAO FAOSTAT Database. Available online: http://www.fao.org/faostat/en/$\delimiter"026E30F$#data/EL. (accessed on 1 November 2019).
4. Bereswill, R.; Streloke, M.; Schulz, R. Current-use pesticides in stream water and suspended particles following runoff: Exposure, effects, and mitigation requirements. *Environ. Toxicol. Chem.* **2013**, *32*, 1254–1263. [CrossRef] [PubMed]
5. Stone, W.W.; Gilliom, R.J.; Ryberg, K.R. Pesticides in U.S. Streams and Rivers: Occurrence and trends during 1992–2011. *Environ. Sci. Technol.* **2014**, *48*, 11025–11030. [CrossRef] [PubMed]
6. Van Metre, P.C.; Alvarez, D.A.; Mahler, B.J.; Nowell, L.; Sandstrom, M.; Moran, P. Complex mixtures of Pesticides in Midwest US streams indicated by POCIS time-integrating samplers. *Environ. Pollut.* **2017**, *220*, 431–440. [CrossRef] [PubMed]
7. FIFRA. *Federal Insecticide, Fungicide and Rodenticide act. U.S. Federal Law, Amended in 1972 and 1988*; FIFRA: Washington, DC, USA, 1947.
8. U.S. Environmental Protection Agency. Reregistration and Other Review Programs Predating Pesticide Registration Review. Available online: https://www.epa.gov/pesticide-reevaluation/reregistration-and-other-review-programs-predating-pesticide-registration (accessed on 30 October 2019).
9. U.S. Environmental Protection Agency. *Overview of the Ecological Risk Assessment Process in the Office of Pesticide Programs, US Environmental Protection Agency Endangered and Threatened Species Effects Determinations*; DIANE Publishing: Collingdale, PA, USA, 2004.
10. U.S. Environmental Protection Agency. *Ecological Effects Test Guidelines OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test*; U.S. Environmental Protection Agency: Washington, DC, USA, 2016; pp. 1–19.
11. U.S. Environmental Protection Agency. *Ecological Effects Test Guidelines OCSPP 850.1010: Aquatic Invertebrate Acute Toxicity Test, Freshwater Daphnids*; U.S. Environmental Protection Agency: Washington DC, USA, 2016; pp. 1–17.
12. U.S. Environmental Protection Agency. *Ecological Effects Test Guidelines OCSPP 850.4400: Aquatic Plant Toxicity Test Using Lemna spp.*; U.S. Environmental Protection Agency: Washington, DC, USA, 2012; pp. 1–24.
13. U.S. Environmental Protection Agency. *Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity*; U.S. Environmental Protection Agency: Washington, DC, USA, 2012; pp. 1–28.

14. US EPA Evaluation Guidelines for Ecological Toxicity Data in the Open Literature. Available online: https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/evaluation-guidelines-ecological-toxicity-data-open (accessed on 30 March 2018).

15. Wolfram, J.; Stehle, S.; Bub, S.; Petschick, L.L.; Schulz, R. Meta-analysis of insecticides in United States surface waters: Status and future implications. *Environ. Sci. Technol.* **2018**, *52*, 14452–14460. [CrossRef]

16. Stehle, S.; Schulz, R. Agricultural insecticides threaten surface waters at the global scale. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5750–5755. [CrossRef] [PubMed]

17. Schulz, R.; Liess, M. A field study of the effects of agriculturally derived insecticide input on stream macroinvertebrate dynamics. *Aquat. Toxicol.* **1999**, *46*, 155–176. [CrossRef]

18. Schäfer, R.B.; Bundschuh, M.; Rouch, D.A.; Szöcs, E.; von der Ohe, P.C.; Pettigrove, V.; Schulz, R.; Nugegoda, D.; Kefford, B.J. Effects of pesticide toxicity, salinity and other environmental variables on selected ecosystem functions in streams and the relevance for ecosystem services. *Sci. Total Environ.* **2012**, *415*, 69–78. [CrossRef]

19. Schäfer, R.B.; Caquet, T.; Siimes, K.; Mueller, R.; Lagadic, L.; Liess, M. Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe. *Sci. Total Environ.* **2007**, *382*, 272–285. [CrossRef]

20. Beketov, M.A.; Kefford, B.J.; Schäfer, R.B.; Liess, M. Pesticides reduce regional biodiversity of stream invertebrates. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11039–11043. [CrossRef] [PubMed]

21. Schäfer, R.B.; Pettigrove, V.; Rose, G.; Allinson, G.; Wightwick, A.; von der Ohe, P.C.; Shimeta, J.; Kühne, R.; Kefford, B.J. Effects of pesticides monitored with three sampling methods in 24 sites on macroinvertebrates and microorganisms. *Environ. Sci. Technol.* **2011**, *45*, 1665–1672. [CrossRef] [PubMed]

22. Vijver, M.G.; Hunting, E.R.; Nederstigt, T.A.; Tamis, W.L.; van den Brink, P.J.; van Bodegom, P.M. Postregistration monitoring of pesticides is urgently required to protect ecosystems. *Environ. Toxicol. Chem.* **2017**, *36*, 860–865. [CrossRef] [PubMed]

23. Malaj, E.; Peter, C.; Grote, M.; Kühne, R.; Mondy, C.P.; Usseglio-Polatera, P.; Brack, W.; Schäfer, R.B. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9549–9554. [CrossRef]

24. Morrissey, C.A.; Mineau, P.; Devries, J.H.; Sanchez-Bayo, F.; Liess, M.; Cavallaro, M.C.; Liber, K. Neonicotinoid contamination of global surface waters and associated risk to aquatic invertebrates: A review. *Environ. Int.* **2015**, *74*, 291–303. [CrossRef]

25. Cairns, J. The myth of the most sensitive species. *BioScience* **1986**, *36*, 670–672. [CrossRef]

26. U.S. Environmental Protection Agency. Pesticides Chemical Search | Chemical Search | Pesticides | US EPA. Available online: https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1 (accessed on 2 March 2018).

27. U.S. Environmental Protection Agency. US EPA's Web Archive: Pesticide Reregistration Status. Available online: https://archive.epa.gov/pesticides/reregistration/web/html/status.html (accessed on 2 March 2018).

28. U.S. Environmental Protection Agency. Aquatic Life Benchmarks and Ecological Risk Assessments for Registered Pesticides. Available online: https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/aquatic-life-benchmarks-and-ecological-risk (accessed on 31 October 2019).

29. U.S. Environmental Protection Agency. *Special Docket for Pesticide Reregistration Risk Assessments-Risk Assessments and Related Documents (DVD Set)*; U.S. Environmental Protection Agency: Washington, DC, USA, 2007.

30. U.S. Environmental Protection Agency. Registration Review Process. Available online: https://www.epa.gov/pesticide-reevaluation/registration-review-process (accessed on 23 October 2019).

31. U.S. Environmental Protection Agency. (2019) ECOTOX User Guide: ECOTOXicology Knowledgebase System. Version 5.0. Available online: http://www.epa.gov/ecotox/2019 (accessed on 1 November 2019).

32. U.S. Environmental Protection Agency. (2019) USEPA Pesticide Ecotoxicity Database. Available online: https://ecotox.ipmcenters.org/index.cfm?menuid=7 (accessed on 1 November 2019).

33. U.S. Environmental Protection Agency. *Ecological Effects Test Guidelines OCSPP 850.1020: Gammarid Amphipod Acute Toxicity Test*; U.S. Environmental Protection Agency: Washington, DC, USA, 2016; pp. 1–18.

34. Beasley, A.; Belanger, S.E.; Otter, R.R. Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way: Stepwise Information-Filtering Tool (SIFT). *Environ. Toxicol. Chem.* **2015**, *34*, 1436–1442. [CrossRef]

35. Connors, K.A.; Beasley, A.; Barron, M.G.; Belanger, S.E.; Bonnell, M.; Brill, J.L.; de Zwart, D.; Kienzler, A.; Krailler, J.; Otter, R.; et al. Creation of a curated aquatic toxicology database: EnviroTox. *Environ. Toxicol. Chem.* **2019**, *38*, 1062–1073. [CrossRef]

36. Bub, S.; Wolfram, J.; Stehle, S.; Petschick, L.L.; Schulz, R. Graphing ecotoxicology: The MAGIC Graph for linking environmental data on chemicals. *Data* **2019**, *4*, 34. [CrossRef]
37. U.S. Environmental Protection Agency. *Environmental Fate and Ecological Risk Assessment Problem Formulation. In Support of Registration Review of Formetanate HCl*; U.S. Environmental Protection Agency: Washington, DC, USA, 2010.
38. R Core Team. *A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2012*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
39. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-98140-6.
40. RPostgreSQL: R Interface to the 'PostgreSQL' Database System. Available online: https://cran.r-project.org/web/packages/RPostgreSQL/index.html (accessed on 1 November 2019).
41. WoRMS. *Editorial Board World Register of Marine Species*; WoRMS: Ostend, Belgium, 2019.
42. Stephen, C.E.; Mount, D.I.; Hansen, D.J.; Gentile, J.R.; Chapman, G.A.; Brungs, W.A. *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*; U.S. Environmental Protection Agency: Washington, DC, USA, 1985.
43. Guiry, M.D.; Guiry, G.M. AlgaeBase. World-Wide Electronic Publication, National University of Ireland, Galway. Available online: https://www.algaebase.org (accessed on 10 August 2019).
44. Pesticide Use in U.S. Crop Production: 2002. Available online: https://pdfs.semanticscholar.org/739e/6ad4b29880c43ab7e8509e0af53111c69735.pdf (accessed on 1 November 2019).
45. Haukoos, J.S.; Lewis, R.J. Advanced statistics: Bootstrapping confidence intervals for statistics with "difficult" distributions. *Acad. Emerg. Med.* **2005**, *12*, 360–365. [CrossRef]
46. Wood, M. Bootstrapped confidence intervals as an approach to statistical inference. *Organ. Res. Methods* **2005**, *8*, 454–470. [CrossRef]
47. Cumming, G. Inference by eye: Reading the overlap of independent confidence intervals. *Stat. Med.* **2009**, *28*, 205–220. [CrossRef] [PubMed]