Institute for Environmental Sciences,
Faculty 7: Natural and Environmental Sciences
University of Koblenz-Landau,
Fortstrasse 7, D-76829
Landau, Germany

RTLE - A New Model to Predict Regulatory Threshold Level Equivalents for Pesticides from the Open Literature

Using the US EPA ECOTOX Database for Large-Scale Environmental Risk Analysis of Pesticides in Surface Waters

– MASTER THESIS –

presented to the
Institute for Environmental Sciences,
University of Koblenz-Landau

Master Program Ecotoxicology

Author
Lara Luisa Petschick
215200753

Supervisors
Prof. Dr. Ralf Schulz
Dr. Sebastian Stehle

April 2018

Selbstständigkeitserklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe und die Arbeit von mir vorher nicht in einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD-ROM).

Landau, 18. April 2018

Abstract

An extensive evaluation of the current state of pesticide contamination of US surface waters was yet restricted as regulatory threshold levels (RTL_{reg}s) are not easily accessible. Therefore, the present study introduces and validates the new RTLE model, that estimates RTL equivalents (RTL_e) for insecticides from data provided in the freely accessible US EPA ECOTOX database. The RTLE model consists out of two consecutive parts: (1) Filter criteria in compliance with the EPA's guideline for the review of open literature were applied to exclude endpoints of low reliability. (2) The most sensitive endpoint among the remaining endpoints per substance is used for RTL equivalent (RTL_e) estimations. Two sets of freshwater (fw) and estuarine (est) RTL_{reg}s for calibration and validation were compiled from US EPA regulatory documents for 83 insecticides. Model calibration aimed at formulating filter criteria such that endpoints were excluded if they were lower than the regulatory relevant endpoint from which RTL_{reg}s were computed. In contrast, the estimation of RTL_es which would overestimate the actual RTL_{reg}s were rather accepted, though it was aimed at minimizing the proportion of over- and underestimations. Calibration of the RTLE model resulted in 18 filter criteria which were applied in addition to a basic query (null model). For validation, an independent set of insecticides (n = 19, and n = 14 for fw and est model applications, respectively) proofed that the RTLE model estimates RTL_es with the same accuracy as for the calibration data (n = 53, and n = 48 for fw and est model applications, respectively). In the fw model, 12.5% (bootstrapped 95% CI 5.6 - 20.8) of RTL_es underestimated the RTL_{reg}, 30.6% (bootstrapped 95% CI 20.8 - 41.7) of estimates were correct, and 56.9% (bootstrapped 95% CI 45.8 - 68.1) of RTL_es overestimated the RTL_{reg}. The est model performed slightly better, as 43.5%(bootstrapped 95% CI 32.3 - 54.8) of RTL_es met the RTL_{reg}, and 40.3%(bootstrapped 95% CI 29.0 - 53.2) of RTL_es overestimated the respective RTL_{reg}, even though the proportion of RTL_es which underestimated the RTL_{reg} was with 16.1%(bootstrapped 95% CI 8.1 - 25.8) higher. Model predictions were significantly improved, if only endpoints were considered for which a regulatory endpoint used for RTL_{reg} calculations was included to the ECOTOX database. However, the likelihood to underestimate the actual RTL_{reg} was not significantly different with regard to the presence or absence of actual regulatory endpoints in the ECOTOX database, such that the model is also applicable for substances which are not listed in the database. The test for dependencies of prediction precision with physical-chemical parameters did not reveal meaningful correlations, for which the model would have to be adapted for. Also, model performance did not rely on a certain sample size of endpoints remaining after filter applications.

The application of the RTLE model to a set of insecticides (n = 29), for which no regulatory documents were available yielded 6 RTL_es. For the remaining 23 substances, no data was available. The test for the applicability of the RTLE model for herbicides (n = 20) and fungicides (n = 15), did not result in a decrease of model precision, even though some minor adaptions were required to maintain the same accuracy as in the original model. Despite the high likelihood to overestimate RTL_{reg}s by the RTLE model application, the more important likelihood to underestimate RTL_{reg}s was low, such that RTL_es do not tend to falsely indicate risks, even though there are non, if compared to environmental concentrations. Overall the RTLE model serves as a promising tool to derive threshold level equivalents and enables a more comprehensive risk evaluation of pesticides in US surface waters.

Contents

1 Introduction

Pesticides are frequently applied to agricultural lands to increase the yield of crops, as their usage intends to control and prevent pest infestations, and the spread of diseases in an intensively managed landscape (see Oerke, 2006, for review). To date, agricultural managed areas comprise a large biome, as in 2015, up to 37% of the world's land area were used for agriculture including arable lands, permanent crops and permanent pastures (FAO, 2017). The worldwide intensification of agriculture is accompanied by an increased use of pesticides, such that a large area of the world's available land is treated with pesticides each year (Matson et al., 1997; Tilman et al., 2001). Once introduced into the environment and since pesticides are potent bioactive molecules, they have the potential to not only impair target organisms, but also unintentionally threaten non-target organisms (Giesy et al., 1999). For instance, if pesticides are applied to adjacent fields, they are likely to enter surface waters (Bereswill et al., 2013; Müller et al., 2002), where there occurrence was linked to alterations in community structures (Liess & Schulz, 1999; Schäfer et al., 2007), impairments of biodiversity (Beketov et al., 2013) and a degradation of ecosystem services on local and global scale (Dudgeon et al., 2006; Millennium Ecosystem Assessment, 2005; Schäfer et al., 2007). In healthy freshwater systems, a variety of species provides important ecosystem services such as nutrient cycling and water treatment (Covich et al., 2004; Vanni, 2002). However, these organisms are vulnerable to the exposure to pesticides and especially insecticides, which mainly act on the neuronal system: Since molecular mechanisms and receptors are evolutionary highly preserved, insecticides are not only likely to treat agricultural pest organisms, but potentially also act on the nervous system of, for instance, freshwater fish and invertebrates (Sanchez-Bayo, 2012). Even though the regulation of pesticides seeks to protect the non-targeted compartments from these undesirable impacts of pesticides (Finizio & Villa, 2002; Pelaez et al., 2013; Touart & Maciorowski, 1997), insecticides (Stehle & Schulz, 2015*a*,*b*), and other pesticides(Szöcs et al., 2017), were shown to threaten surface waters by frequently exceeding regulatory threshold levels (RTL_{reg}s). Therefore, further efforts are essential to improve the protectiveness of ecological risk assessment as a basis for a more sustainable regulation and use of pesticides (Stehle & Schulz, 2015*b*). Here, governmental and scientific monitoring programs are indispensable to evaluate the current state of protectiveness. For instance, the US Environmental Protection Agency (EPA) provides guidance and regulatory documents from which regulatory threshold values can be computed, as well as a comprehensive monitoring data set including pesticide concentrations in US surface waters (WQP, 2018). This data is freely accessible, such that it is possible to evaluate the protectiveness of the current state of ERA by comparing monitoring data to RTL_{reg}s. However, until now the extensive analysis of environmental samples was hindered by a limited availability

of regulatory threshold values, as each RTL_{reg} needs to be retrieved from a set of regulatory documents, which is time-consuming. Therefore, this work aims at facilitating the risk analysis of large environmental samples by providing a tool for automated threshold level estimation, the new Regulatory Threshold Level Estimator (RTLE) model.

1.1 Regulatory environmental risk assessment of pesticides in the USA

So far, in the United States, pesticides which are sold and distributed within the US, need to be registered under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA, CFR40 §152; FIFRA, 1947). This registration seeks to ensure that "no unreasonable adverse effects on [...] the environment will occur" (US EPA, 2013) by the use of pesticides, what shall be examined in an ecological risk assessment. In brief, the US EPA performs ecological risk assessments in a three-phase-iterative process, in which a problem formulation is elaborated, followed by a risk analysis and a risk characterization, which summarizes the identified risks and proposes management options (US EPA, 1998). For risk analysis, ecological toxicity and exposure data is required. In general, a measure of exposure (either a measured or an estimated environmental concentration) is divided by a measure of toxicity (e.g., EC_{50} for *Daphnia magna*) to obtain a risk quotient (RQ) as (US EPA, 2015*c*):

$$\mathrm{RQ} = \frac{\text{Exposure}}{\text{Toxicity}} \tag{1}$$

This RQ is then compared to a Level of Concern (LOC, e.g., acute high risk to aquatic animals LOC = 0.5; for aquatic plants LOC = 1.0) which shall not be exceeded and is comparable to applying a safety factor. The relevant endpoints are obtained from registrant-submitted toxicity studies that satisfy the data needs under FIFRA §152. Thereby, the aim is to display the actual state of scientific knowledge by also evaluating the suitability of peer-reviewed open literature to either fill identified data-gaps or to refine RQ calculations if valid, more sensitive endpoints are reported (Brady, 2011).

Open literature ecological toxicity data in EPA pesticide risk assessment

To satisfy EPA's data requirements for the registration of pesticides as stated under FIFRA CFR40 §152, ecological effects data from the open literature (i.e., predominantly peer-reviewed studies, but also grey literature as agency publications; US EPA, 2017*c*, p. 46) needs to be evaluated, and comparable studies to guideline studies need to be considered in ecological risk assessment performed by the Office of Pesticide Programm (OPP). The review process includes a first screening followed by a comprehensive validity evaluation of potentially relevant

open literature studies to check whether study results are robust and suitable for quantitative or qualitative use in risk assessment and risk characterization (Brady, 2011). If endpoints provided in peer-reviewed studies are more sensitive (i.e., lower reported effect concentrations, higher toxicity) than endpoints from registrant-submitted standard tests, open literature primary studies (i.e., published in peer-reviewed journals, original studies, no reviews) need to be considered in ecological risk assessment.

The open literature relevant to ecological risk assessments performed by the EPA is searched and stored in its open literature database, the ECOTOX search engine (https://cfpub.epa.gov/ecotox/index.html). The EPA ECOTOX database includes effect toxicity data for whole organisms in aquatic and terrestrial compartments for single chemicals with known CAS number, which are likely to be released into the environment (e.g., pesticides). Since the database is accessible online for the public, it is possible to retrieve ecological effects data of studies which are potentially included to the review process of open literature within the elaboration of regulatory documents as Registration Reviews for pesticides. To be encoded into the ECOTOX database, studies need to fulfill five minimum criteria, and nine further OPP criteria need to be passed to be considered in the review of open peer-reviewed literature by EPA's risk assessors. Thus, all studies referenced in the ECOTOX database have passed the five minimum criteria (box 1), and studies which are relevant to OPP risk assessors fulfill at least the nine additional criteria (box 2) to be included to ecological risk assessments.

Box 1: ECOTOX minimum criteria according to Brady (2011):

1. The toxic effects are related to single chemical exposure;
2. The toxic effects are on an aquatic or terrestrial plant or animal species;
3. There is a biological effect on live, whole organisms;
4. A concurrent environmental chemical concentration/dose or application rate is reported; and
5. There is an explicit duration of exposure.

For studies which pass the above quoted criteria (box 1 and 2), features as test organisms, type of medium, test condition, and reported endpoints are encoded into the database. The screening of open literature is followed by a detailed literature review. Within this review process of the open literature, studies are evaluated for their suitability and reliability in ecological risk assessment by applying further validity criteria (e.g., attachement 5 to Brady, 2011). To be considered in further risk evaluations, studies need to be classified as *scientifically valid*. Valid studies are again distinguished by OPP's risk assessors to be either suitable for quantitative use,

e.g., can be used in risk quotient calculations, or studies which are only suitable for qualitative use, e.g., are considered in risk descriptions, but are not suitable for calculations. All studies categorized as suitable for quantitative risk quotient calculations need to pass approximately the same criteria as the guideline studies submitted by the registrants (Brady, 2011; for comparison of standard guideline criteria, please refer to A3.1 and A3.2, appendix). If studies do not pass the screening or evaluation of validity criteria, they are categorized as *invalid*, and not deemed relevant or defensible. Consequently, invalid studies are excluded from further analysis and neither used for quantitative nor qualitative risk evaluations.

Box 2: Additional OPP criteria according to Brady (2011):

6. Toxicology information is reported for a chemical of concern to OPP;
7. The article is published in the English language;
8. The study is presented as a full article;
9. The paper is a publicly available document;
10. The paper is the primary source of the data
11. A calculated endpoint is reported;
12. Treatment(s) are compared to an acceptable control;
13. The location of the study (e.g., laboratory vs. field) is reported; and
14. The tested species is reported and verified.

To sum up, the EPA ECOTOX database provides ecological effects data from the open literature which might be used in Registration Reviews as long as studies are in compliance with the standard guidelines published by the EPA's Office of Chemical Safety and Pollution Prevention (OCSPP) and if a higher toxicity than in registrant-submitted studies is indicated. The encoding of study features into the ECOTOX database allows to filter for certain standard validity criteria to estimate which endpoints might be relevant for risk assessment. Therefore the ECOTOX database is potentially a suitable tool to retrieve ecological effects data for chemicals of interest in the present study.

1.2 Scientific environmental risk assessment of pesticides in surface water bodies

As one part of regulatory environmental risk assessment of pesticides, threshold values are compared to environmental concentrations of these pesticides measured or predicted in surface waters. In scientific risk evaluations, to determine the risk for non-target organisms and the current degree of protectiveness of pesticide regulation, toxicity endpoints can be compared

to the actual or predicted exposure in surface waters by calculating regulatory threshold levels (Stehle & Schulz, 2015*a*). This risk evaluation follows the same principles of the regulatory risk assessments with the only difference that the measure of toxicity is first multiplied with the LOC and then compared to measures of exposure, as

$$\mathrm{RTL_{reg}} = \mathrm{Toxicity} \times \mathrm{LOC} \tag{2}$$

To test for instance, the protectiveness of regulatory risk assessments, regulatory threshold levels are computed by multiplying the most sensitive endpoint from a valid study reported in the respective regulatory document with a Level of Concern. Then, to assess the risk of pesticides, $\mathrm{RTL_{reg}}$s are compared to actual (e.g., in monitoring studies) or predicted (e.g., in modeling studies) environmental concentrations. If the $\mathrm{RTL_{reg}}$ is exceeded, it is assumed that the ecosystem is not sufficiently protected and detrimental effects might occur.

Up to now, these endpoints were retrieved by reviewing the most recent regulatory document, identifying the most sensitive, valid endpoint and applying the respective LOC for each substance, which is time consuming. Additionally, EPA's search engines need to be checked frequently for new regulatory document publications to keep the list of $\mathrm{RTL_{reg}}$s up-to-date. Unfortunately, regulatory documents are not provided on a single web page, but need to be searched from four different sources always looking for the most recent document and the lowest reported endpoint. Further, not all regulatory documents contain the same information. Some *Reregistration Eligibility Decisions*, for instance, contain qualitative risk characterizations only but no effect data, whereas others do report the required study endpoints and also report risk quotient calculations. Therefore, per substance, a multitude of documents needs to be downloaded and reviewed until the respective valid, most sensitive endpoint is retrieved and confirmed which takes from hours to days. Thus, a comprehensive analysis of the pesticides' impact on the biotic environment was yet limited to the amount of completed reviews of regulatory documents, since governmental monitoring datasets (e.g., WQP, 2018) comprise a large number of different pesticides, but a comprehensive list of $\mathrm{RTL_{reg}}$s is missing. To enable an extensive estimation of the current state of exposure of non-target organisms to pesticides, it therefore was aimed at automating the process of endpoint review using the available data provided in the EPA ECOTOX database.

1.3 Study objectives and model assumptions

Since measured environmental concentrations of pesticides are seldom related to regulatory threshold values for risk evaluations in the current literature and an extensive analysis with re-

gards to the number of considered pesticides is lacking, the knowledge on the degree of pollution of surface waters by pesticides and especially insecticides is yet limited. The new RTLE model aims at enabling and facilitating the scientific evaluation of insecticide exposure in the field by providing threshold level estimates for an enlarged number of substances. Therefore, the aim of the present study was to predict Regulatory Threshold Level equivalents (RTL_e) for insecticides from the open literature provided in the EPA ECOTOX database. The RTLE model was set up to (a) simulate the evaluation of open literature for suitability to be used in ecological risk assessments within the Registration Review process of insecticides (yellow path in fig. 1), and (b) to compute RTL_es from relevant endpoints which passed the validity screening. Thereby, it was aimed at approximating the respective endpoint within the set of reported endpoints per substance for which threshold estimations meet the RTL_{reg}. The closer the estimate gets to the RTL_{reg}, the higher is the model precision.

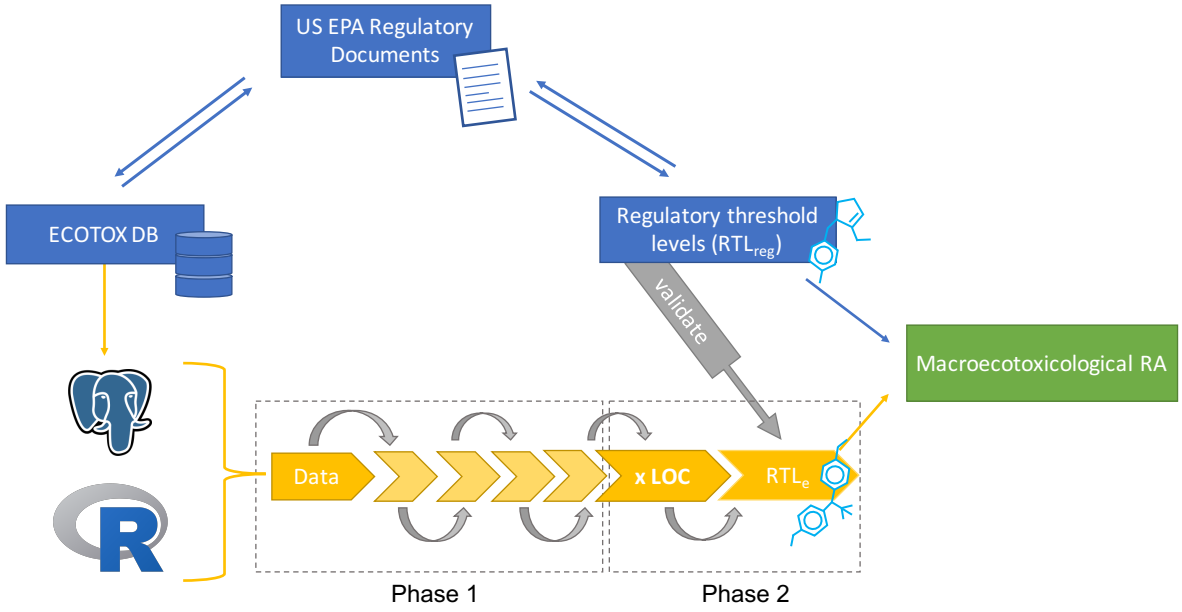


Figure 1: Schematic overview on RTL_{reg} and RTL_e generation for risk assessment of insecticides in surface waters. Yellow path indicates the workflow of the RTLE model: first, selected filter criteria are used to exclude irrelevant data (model phase 1) to predict RTL_es by applying the LOC to the selected relevant endpoint (model phase 2). Regulatory documents were searched for relevant endpoints and RTL_{reg}s were computed. The RTLE model was validated by comparing RTL_{reg}s to RTL_es thereby assessing the model precision of predictions. Resulting RTL_es could be used for macroecotoxicological risk assessment.

The RTLE model consists out of two parts: The first is a database query, which mimics the validity check of open literature according to EPA standard guidelines (Brady, 2011). The second part computes RTL_es from the selected most sensitive endpoint which has passed the screening process by applying the respective LOC (fig. 1, phase 1 and 2). Thereby, the model relies on the following assumptions: (1) the most sensitive endpoint is relevant for RTL_e computations, (2) acute endpoints are representative to mimic insecticide exposure in the field, and (3) aquatic

plant endpoints can be excluded from model predictions. These assumptions are motivated as follows: It was hypothesized that first, the lowest endpoint which passes the validity screening by the applied filter steps would approximate the endpoint that would be chosen for risk quotient calculations. Assuming that the endpoints relevant to risk assessment are available in the ECOTOX database, the endpoints remaining after filter application should be relevant for threshold generation. Second, insecticide concentrations in surface waters occur in short peaks such that no chronic exposure is expected (Stehle et al., 2013) and it is justifiable to consider only acute toxicity data for RTL_e estimation. Third, acute aquatic plant endpoints do not need to be considered for insecticide RTL_e estimations, since aquatic plants are in most cases not the most sensitive species if exposed to insecticides, as indicated in the reviewed regulatory documents (table A3 and A4, respective endpoint types for validation and calibration datasets). To test the performance of the model, predictions (RTL_es) are compared to actual regulatory threshold levels and the likelihood to over- or underpredict these threshold levels is estimated. The model performance is further evaluated by calculating *Pearson's correlation coefficients* to quantify the degree to which prediction precision is associated with physical and chemical parameters: The solubility, partitioning coefficient and molecular weight of substances were chosen to be compared to a measure of model precision. These three parameters are known to be related to the toxicity of a substance (Gama et al., 2012). Hence, the test for relations was performed to check whether patterns would translate from the measure of toxicities (threshold values) into the measure of model precision. Further, it is tested, whether there are insecticide groups (e.g., pyrethroids, organophosphates) or a type of datasets used (i.e., grouping into calibration and validation datasets according to their occurrence in the Water Quality Portal (WQP) dataset and pesticide sales data (Gianessi & Reigner, 2006; WQP, 2018)) for which the model performs better or worse. In the end, it is tested whether the model could be extended to predict RTL_es for additional pesticide groups (i.e., herbicides and fungicides).

2 Materials and Methods

2.1 Review of regulatory documents and regulatory threshold value generation

To obtain a list of acute aquatic toxicity endpoints of insecticides used for environmental risk assessment, the most recent regulatory documents were searched online and reviewed. Documents provided via the EPA Pesticide Chemical Search engine (US EPA, 2017*d*), Aquatic Life Benchmarks (US EPA, 2017*a*), the EPA archive (US EPA, 2017*e*), and the Special Dockets DVD content (US EPA, 2007) were reviewed and ecological effects data was compiled from the most recent available documents. Thereby, pesticides were searched by either CAS number

or name to allow an unambiguous identification of each test substance. A list of relevant substances was compiled from the WQP dataset which provides, aside others, data of pesticide monitoring in the United States (WQP, 2018). For model calibration and validation, documents for insecticides which occur regularly in the WQP dataset were reviewed. To increase the number of substances for model validation, the WQP-list of substances was complemented with a list of substances from an insecticide sales report (Gianessi & Reigner, 2006). Subsequently, to test whether the model is applicable to further pesticide groups, herbicide and fungicide data was retrieved from regulatory documents for a set of randomly selected substances occurring in the WQP dataset to obtain an extension dataset. Regulatory documents were available for 59 substances comprising the calibration dataset, for 24 substances comprising the validation dataset, and for 15 fungicides, and 27 herbicides (extension dataset). For all substances, acute toxicity endpoints for freshwater, and estuarine test organisms were recorded from the reviewed documents. If more than one endpoint per species group was reported, the lowest, most sensitive endpoint was chosen, or preferably the endpoint, which was used to calculate risk quotients, whenever RQ calculations were conducted and sufficiently reported. Regulatory threshold levels were calculated from the most sensitive acute aquatic toxicity endpoint (i.e., EC_{50}, LC_{50} or IC_{50}) for the different compartments freshwater and estuaries, to allow the estimation of freshwater RTL_es and estuarine RTL_es. The US regulatory LOC was used to compute regulatory threshold values according to formula 2 introduced in part 1.2 of the present study. As defined by the US EPA, the LOC for acute fish and invertebrate studies is 0.5. The LOC for aquatic plant studies considered in the extension data set is 1.

2.2 Species selection

A list of standard test species was compiled from OCSPP acute ecological effects guidance documents (OCSPP 850.1735, 850.1075, 850.1055, 850.1045, 850.1035, 850.1025, 850.1020, 850.1010, US EPA (2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)). Further, OECD guidelines were used to extract additional chironomid and daphnia species used in standard tests (OECD, 2004*a*,*b*, 2011). For the model extension, aquatic plant species were included (OCSPP 850.4400, 850.4500, 850.4550, (US EPA, 2012*a*,*b*,*c*)). Finally, further species were deemed relevant for risk assessment which previously were selected as endpoints in regulatory documents. A list of selected model species ist provided in the appendix (table A5).

2.3 EPA ECOTOX database

The most recent version of the whole ECOTOX database was downloaded (https://cfpub.epa.gov/ecotox/, updated by 12/14/2017) and a PostgreSQL local library was built

to store the available ecotoxicological data (following the suggested workflow introduced by Szöcs (2016)). Further, a table with standard test species was imported into the database (table A5) as well as two tables including the calculated RTL_{reg}s, and information on the publication year of the most recent regulatory documents for estuarine and freshwater endpoints, respectively (table A3 & A4). The database was built and queried with the help of R (R Core Team, 2016) and the RPostgreSQL package (Conway et al., 2016), as well as PostgreSQL 9.5.3 and pgAdmin 4 2.0. All calculations were performed in R (Dowle & Srinivasan, 2017; Pruim et al., 2017). For graphics, the package ggplot2 by Wickham (2009) with extensions (Auguie, 2017), and the package car (Fox & Weisberg, 2011) were used.

2.4 Filter criteria selection and model calibration

In general, the selection process of relevant endpoints aimed at simulating EPA's standard procedure for review of open literature of ecological toxicity data to be used in US pesticide registrations (Brady, 2011), and the validity criteria stated in OCSPP standard test guidelines (US EPA, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*). If possible, validity criteria were translated into SQL code and the local version of the US ECOTOX database was queried to filter relevant endpoints. Thereby, filter criteria were formulated to exclude datapoints (e.g., exclusion of studies which report multiple effects in one endpoint, exclusion of studies with organisms of unreliable sources) that would be rated as *invalid* or *for qualitative use only* during the OPP's review of open literature for pesticide registration. datapoints excluded by the SQL filter criteria application are deemed irrelevant, whereas datapoints that pass the filter are likely to be relevant for RTL_e calculations. During the process of filter criteria selection, the SQL query was adapted in an iterative manner to improve filter performance but avoiding the exclusion of relevant data by too strict filter criteria applications to keep a sufficient set of data and substances.

To calibrate the model, the best performing combination of filter criteria was searched. The calibration of the model seeks at minimizing the proportion of substances for which threshold estimates would be too low while at the same time it was aimed at minimizing the breadth of the scatter in the data. After filter criteria were defined and applied, their influence was evaluated by comparing the proportion of datapoints excluded by single filter step applications and additive filter applications. Filter criteria which did not lead to an improvement of filter outcome were removed from the query. Then, from the set of relevant filter criteria, a null model, a mid model and a full model were formulated, which differed in their strictness of validity criteria, whereby the null model contained the most simple set of filter criteria being least restrictive, the mid model contained more filter criteria, but was not as highly specific as the full model, which contained the largest number of filter criteria.

To evaluate the performance of model application, ratios between each remaining endpoint after applying the filter criteria and the actual regulatory endpoint were calculated. This ratio was used as a measure of model precision: If threshold estimates met or exceeded the actual RTL_{reg}, $\log_{10}\left(\frac{RTLe_i}{RTLreg_i}\right) \geq 0$. If threshold estimates undershot the actual RTL_{reg}, $\log_{10}\left(\frac{RTLe_i}{RTLreg_i}\right) < 0$ for each endpoint i. While for model calibration, the undershooting of the actual RTL_{reg} was avoided or kept as low as possible, since it would result in too low RTL_es and falsely indicate a high toxicity of a substance resulting in an overestimation of the risk in environmental risk evaluations, the overshooting of the RTL_{reg} was rather accepted, since if RTL_e estimates exceeded the actual RTL_{reg}, they would rather underestimate the true risk if used in risk evaluations. An underestimation was rather accepted, since if the partly overestimated threshold equivalents exceed environmental samples, it is likely that there is indeed a risk. Vice versa, if threshold estimates would be too low, it could not be excluded, that there is no risk, even though a risk was indicated by environmental concentrations exceeding the predicted threshold levels. Therefore, the model calibration aimed at finding the best filter combination to maximize the proportion of insecticides for which the ratio $\log_{10}(RTL_e / RTL_{reg})$ approached zero. And, to not falsely overestimate the risk of pesticides to non-target organisms, it was aimed at excluding the more sensitive endpoints by applying additional filter criteria, if studies were deemed not relevant for governmental risk assessment due to invalidity or low reliability of data source.

An exclusion rate analysis was performed for single and additive filter applications, respectively. To test the influence of single filter criteria, the proportion of remaining data points was compared to the most basic model, the null model, for each applied filter. Filter criteria which did not exclude any irrelevant data points were removed from the SQL code. Then, to test the influence of additive filter applications, the distribution of ratios ($\log_{10}(RTL_e / RTL_{reg})$) among all insecticides was tracked for each filter application. Subsequently, the proportions of the most sensitive endpoints (minimum) among all insecticides lying below, at, or above zero were evaluated, as well as the number of insecticides, which remained after each filter step application.

After filter refinements, the position of the endpoint leading to the respective RTL_{reg} within the distributions of filtered datapoints was determined per substance. Then, the median quantile of the regulatory relevant endpoint positions among all insecticides as well as 95% confidence intervals were calculated for each model query (null, mid and full, as described by Campbell & Gardner, 2005). These quantiles were then used to estimate the most likely (median) regulatory endpoint position within the distribution of endpoints. The estimated position was then used to extrapolate the approximate RTL_{reg} position within the distribution of datapoints for further insecticides with unknown RTL_{reg} as RTL_e.

When relevant filter criteria were selected and the position of the corresponding regulatory endpoint was determined, the second part of the model was formulated such that RTL_es were computed from relevant endpoints by applying the respective LOC. This part was especially relevant for model extension, since for the risk estimation to aquatic plants, different LOCs are used than to assess the acute risks to aquatic animals.

2.5 Model validation

The exactness of RTL predictions were validated by reviewing the regulatory documents of 24 further insecticides (validation dataset) to retrieve their most sensitive regulatory endpoints and calculate RTL_{reg}s. Then, results from predictions were compared to the actual regulatory endpoints and potential differences in the predictions for calibration and validation datasets were evaluated by hypothesis testing. If the underlying assumptions (normal distribution of the data and homoscedasticity) of the Student's t-test were violated after data transformation, a non-parametric version, the Wilcoxon Rank Sum test, was used to analyze potential differences in predictions from the calibration and validation datasets. Then, the performance of predictions for included versus non-included endpoints in the ECOTOX database was evaluated. For this analysis, the model query was adapted to compile only data for substances, for which an endpoint is reported, that leads to the actual RTL_{reg}, if the LOC was applied. To test the influence on the precision of threshold predictions of whether datapoints were included or not included in the ECOTOX database, again (parametric or non-parametric) hypothesis testing was applied.

To account for potential difficulties in prediction power or interrelations, it was checked, whether there were groups of substances which were more difficult to predict or any correlations to physical-chemical parameters: Pearson's correlation moments were calculated between the K_{OC}, molecular weight of substances (MW), as well as the level of solubility, and the ratio $\log_{10}(RTL_e/RTL_{reg})$, RTL_{reg} and RTL_e, respectively, using all available data by pooling the calibration and validation datasets. Data of physical and chemical parameters of substances were retrieved from the Pesticide Action Network (PAN, Kegley et al., 2016) database, which enabled the analysis of 54 and 49 substances for freshwater and estuarine model predictions, respectively. Additionally, it was analyzed whether the predictions were relying on a certain sample size for each substance to be reliable, or whether there were substance groups, for which predictions performed better or worse. Here, the pooled calibration and validation data was applied (n = 72 for freshwater model predictions, n = 62 for estuarine freshwater model predictions).

Assuming that the sample of retrieved RTL_{reg}s of the calibration and validation data sets are representative for all insecticides in the ECOTOX database, the likelihood to over- and under-

estimate the risk by using RTL_e in scientific risk assessments was evaluated. Therefore, mean proportions with bootstrapped confidence levels were calculated. Samples of the same size as the original datasets were bootstrapped with replacements (Campbell et al., 2005). From each distribution, the proportions of $\log_{10}(RTL_e/RTL_{reg})$ exceeding zero, equaling zero or being lower than zero were computed. Samples were drawn 1000 times and the mean was calculated for each generated proportion. 95% confidence intervals were estimated from the distribution of data by taking quantiles (e.g., 2.5% and 97.5% centiles) of the bootstrapped distributions as described by Campbell et al. (2005). This enabled the statistical comparison of groups by checking whether confidence levels were non-overlapping or partly overlapping (if less than half of the confidence intervals overlapped), which would indicate a significant difference between two groups with $p<0.01$ or $p<0.05$, respectively (Cumming & Finch, 2005). Finally, the influence of whether datapoints were included in the database or not on the likelihood to over- or underpredict RTL_es applying the full model, was tested by bootstrapping means and confidence intervals as described above and results were compared to the full model performance of predictions.

2.6 Model application

First, the best performing model was applied to retrieve RTL_es for a set of insecticides, for which no regulatory documents could be found (n = 29). Then, it was checked whether the model could be applied for herbicides and fungicides, two further groups of pesticides. It was evaluated, if additional filter criteria would need to be applied to improve the performance of model predictions for fungicides and herbicides. Therefore, regulatory documents were reviewed for a set of fungicides and herbicides, whereby the SQL query was adapted to no longer exclude aquatic plant endpoints. To evaluate the extended model's performance, (parametric or nonparametric) hypothesis testing was used to compare model precision of the original full model and the extended full model. In more detail, the results from predictions of the following groups were compared: insecticide prediction precision (for validation and calibration substances) was compared to herbicide and fungicide prediction precision, once in the freshwater model and once in the estuarine model.

3 Results

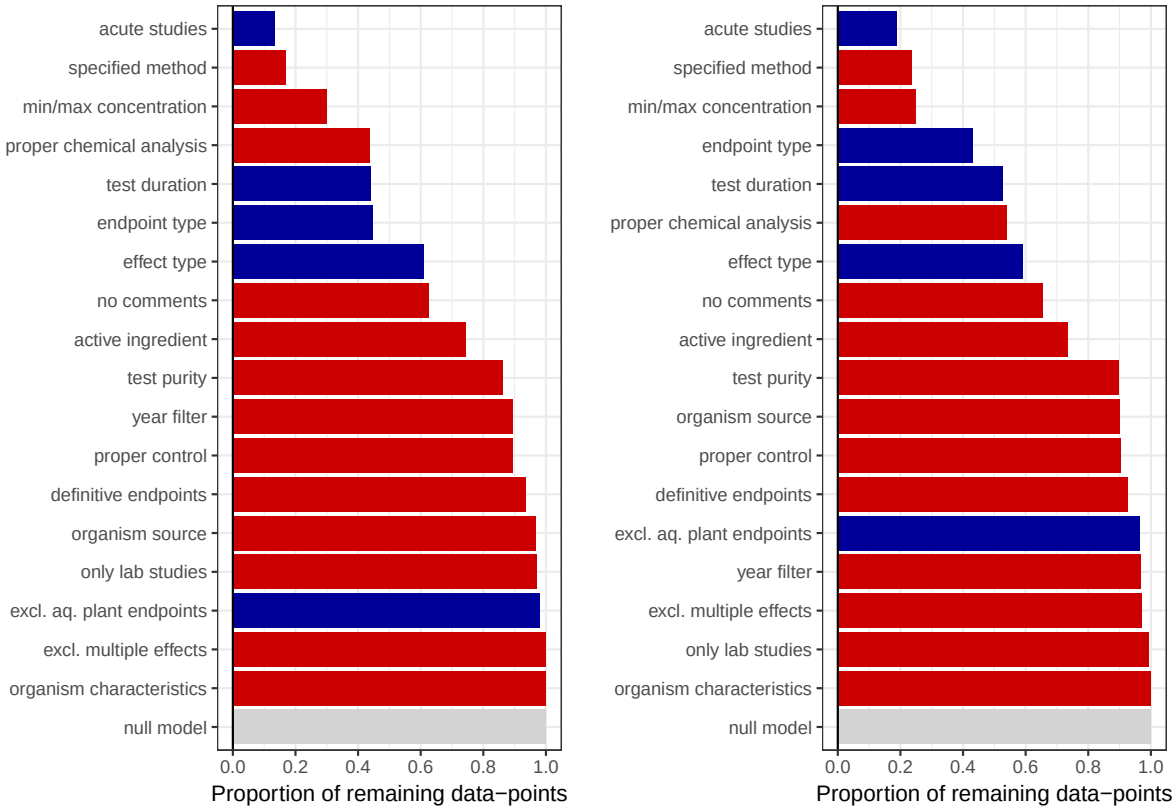
3.1 Model calibration

3.1.1 Filter criteria selection and influence of filter application

From the SQL query adaption, three different models were elaborated, which differed in their strictness of data exclusion, and will be further explained in the following. The iterative filter refinement resulted in 18 filter criteria (table 1) that were applied in addition to the null model to yield the full model. The basic query, the null model, was the least strict model and filtered relevant substances by CAS number, model species, the concentration unit (ug/L, or convertible to ug/L), and the respective type of medium for estuarine (SW: salt water) and freshwater (FW: freshwater) organisms (step 0, table 1). The mid model contained further refinements which filtered for acute study types (respective endpoint, effect type, study duration, acute studies) and excluded aquatic plant endpoints, which were deemed irrelevant since the majority of aquatic plants is rather non-sensitive to insecticides (steps 1-5, table 1). The full model included further refinements, which were adapted iteratively to reduce the spread in the predictions and especially reduce the rate of underestimations of the RTL_{reg}. The application of each filter step was evaluated for freshwater and estuarine models. About half of the filter criteria (9 out of 19) had only small filter power, as they excluded less than 15% of the datapoints in comparison to the null model (fig. 2). The lowest influence had the *exclusion of multiple effects* filter and the *organisms characteristics* filter for the freshwater model and the *lab study* filter and the *organisms characteristics* filter in the estuarine model (fig. 2). The *active ingredient*, *no comments*, and *effect type* filter excluded about 25 - 40% of datapoints in both models. After application of the *endpoint type*, *test duration*, and *proper chemical analysis* filter about 60% of datapoints were excluded in the freshwater model and about 50 - 60% of datapoints in the estuarine model. The mid model which comprise all blue bars (fig. 2) included the filter criteria with the highest exclusion power, whereas most filter criteria formulated for refinements in the full model (red bars) did not exclude a high proportion of datapoints. This was also evident from the analysis of additive filter applications: the mid model contained filter criteria which removed most of the scatter in the data sets pooled among all substances (fig. 3, steps 0-6 for freshwater systems, 0-3 for estuarine systems), whereas the full model filter criteria did not result in large changes in the range of predictions, but helped to exclude some endpoints, which would result in underestimations of the actual regulatory threshold level (fig. 3, approximately steps 7-12 for freshwater systems, and 4-12 for estuarine systems). The addition of further filter criteria (12-18) did not lead to major changes of data scattering in both model systems. The range of endpoints exceeding the RTL_{reg} decreased with the additive application of filter criteria, but overestimated

the RTL_{reg}s by up to 5 to 6.5 orders of magnitude for freshwater and estuarine models, respectively. However, 50% of the data lay within approximately 2 orders of magnitude, and 95% of data within about 4 orders of magnitude (fig. 3, grey bars) for both freshwater and estuarine predictions. Further, for freshwater and estuarine endpoint types, about 5% of ratios lay below 0 applying the full model (i.e., underoverestimateestimated the actual RTL_{reg}) by 0.67 and 0.34 orders of magnitude, respectively (fig. 3).

For freshwater and estuarine data, at the start of additive filter applications, data for 58 and 54 substances were available, respectively (fig. 4). Simultaneously, the majority of remaining datapoints after filter application results in too low threshold estimates (fig. 4, $\log_{10}(RTL_e/RTL_{reg}) < 0$). After 4 (estuarine model application) - 5 (freshwater model application) filter steps, the proportion of substances with data points that were more sensitive than the regulatory endpoint largely decreased with only slight changes ,if further filter steps were applied, such that the following filter steps could be considered fine-tuning, while the number of substances decreases to 53 and 48 substances, for freshwater and estuarine model applications, respectively.



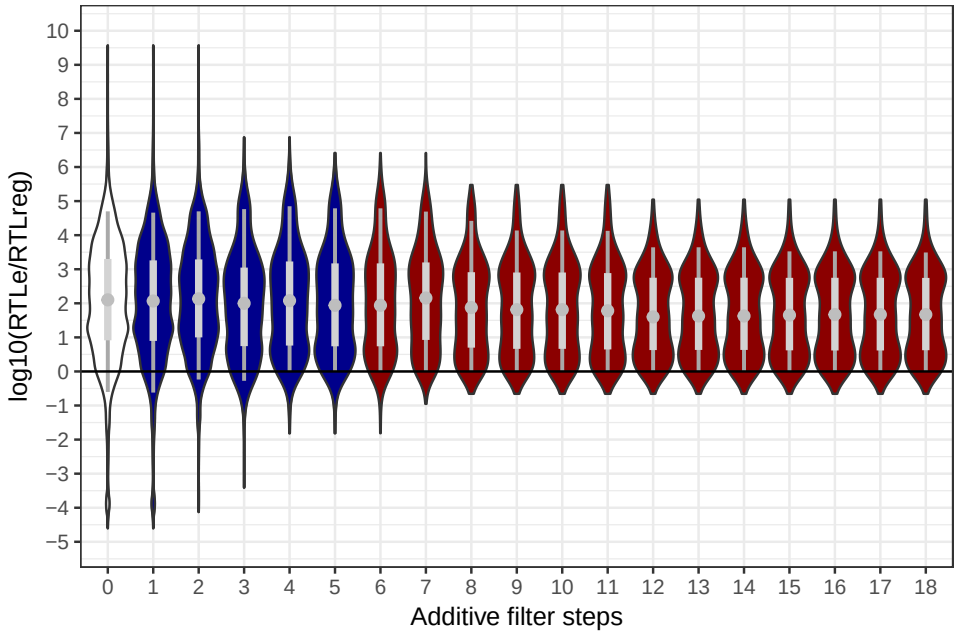
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure 2: Influence of single filter criteria application in relation to the null model. Proportion of remaining datapoints after single filter criteria application for freshwater (a, null model: n = 14581) and estuarine (b, null model: n = 2709) endpoints. Filter criteria are ordered descending according to their exclusion power with most powerful criteria on top of the flipped bar-plot. Grey bars belong to the null model, blue bars belong to the mid model and red bars belong to the full model. Largest proportions of removed datapoints are yielded after application of mid model filter criteria.

Table 1: Filter steps applied to freshwater and estuarine ECOTOX data for the full model, mid model and null model, and criteria necessary for the model extension for herbicides and fungicides (for SQL code, please refer to the appendix A2.5)

NULL MODEL FILTER CRITERIA	
0	Select basic unit which is properly convertable to ug/L, relevant substances by CAS number, and relevant test species via standard test species filter. Filter for relevant medium: – for freshwater organisms: FW = freshwater medium – for estuarine organisms: SW = salt water medium
MID MODEL FILTER CRITERIA	
1	Exclude aquatic plant endpoints for insecticide predictions, since most aquatic plants were not or less sensitive to insecticide exposure in regulatory effect summaries
2	select mortality and intoxication as effect types
3	Select studies with acute or not reported test types. Studies with not reported test types was allowed to avoid exclusion of relevant data points.
4	select relevent endpoints for acute test types, LC_{50}, EC_{50}, IC_{50}
5	select relevant observation times for acute studies: – 2, 3 or 4 days, and 48, 72, 96 hours for animals – 2, 3, 4, 5, 7, 14 days, and 48, 72, 96 hours for plants
FULL MODEL FILTER CRITERIA	
6	Exclude studies which report multiple effects in one endpoint.
7	Exclude studies for which effect comments were given (e.g., toxicity symptoms, aeration).
8	Select only purity above or equal to high percentage (70%). Allow studies with not coded or not reported purities, since otherwise too many endpoints would be excluded.
9	Select studies with active ingredient as concentration type. Exclude not coded and not reported concentration types.
10	Allow only laboratory studies.
11	Exclude references younger than regulatory documents. Filter step applied only for model calibration and validation, not for model application.
12	Exclude studies which do not report minimum and maximum concentrations, which is deemed necessary to generate LC_{50}/EC_{50}/IC_{50}
13	Exclude organisms from unreliable sources – Exclude organisms from the wild – Exclude organisms from not reported sources
14	Exclude tests with non-definitive endpoints, since no reliable dose-response relationship can be provided.
15	Test methods need to be specified. Exclude studies for which test methods are not coded or not reported
16	Exclude tests in which no chemical analysis was performed. Here, studies with not reported and not coded chemical analyses were not excluded, since otherwise deemed relevant datapoints might be lost.
17	Exclude endpoints with insuffcient control: If the control is unsatisfactory, insufficient, outside of the primary exposure system, historical, or no control was specified or no control was used.
18	Exclude test organisms older than 24 hr (Daphnia standard, important for fenamiphos), and fry stages of fish.
EXTENDED MODEL FILTER CRITERIA	
Aq. plant 1	allow plant endpoints
Aq. plant 2	select relevant test duration for plant studies
Aq. plant 3	include aquatic plant effect types (growth, reproduction, population)

(a) Freshwater endpoints



(b) Estuarine endpoints

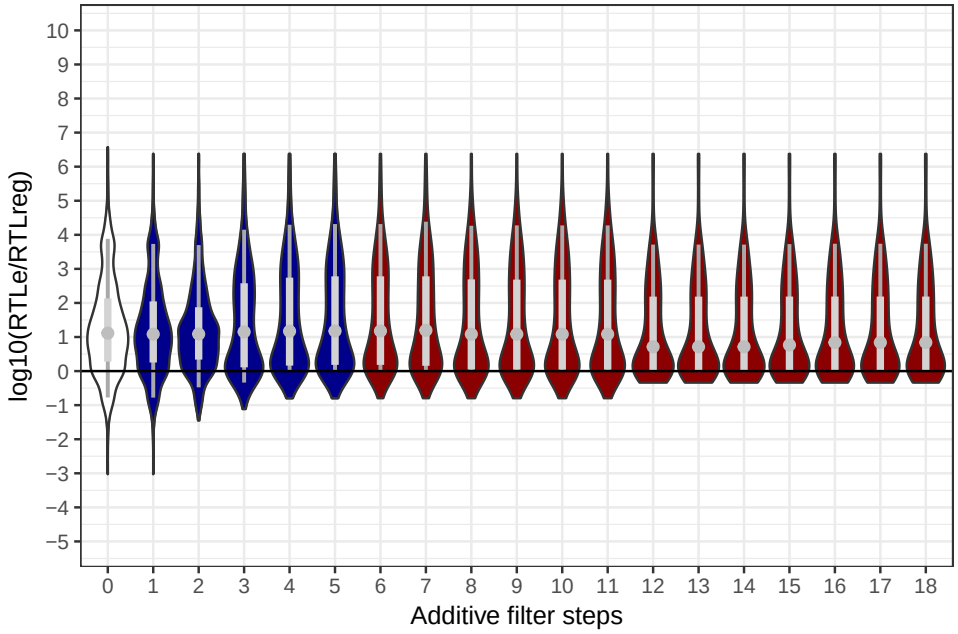


Figure 3: Influence of additive filter step application on proportion of over- and underpredictions of the RTL_{reg} expressed as ratio $\log_{10}(RTL_e/RTL_{reg})$ greater, equal or lower than 0 among all substances for (a) freshwater and (b) estuarine filter applications. Violin colors refer to the applied models: null model: white, mid model: blue, full model: red. Grey dots indicate position of the median, light grey boxes indicate the extent of 50% of datapoints. Dark grey bars indicate extent of 90% of datapoints around the median. Most datapoints are excluded within the first filter steps and the majority of the remaining ratios is close to or exceeds 0.

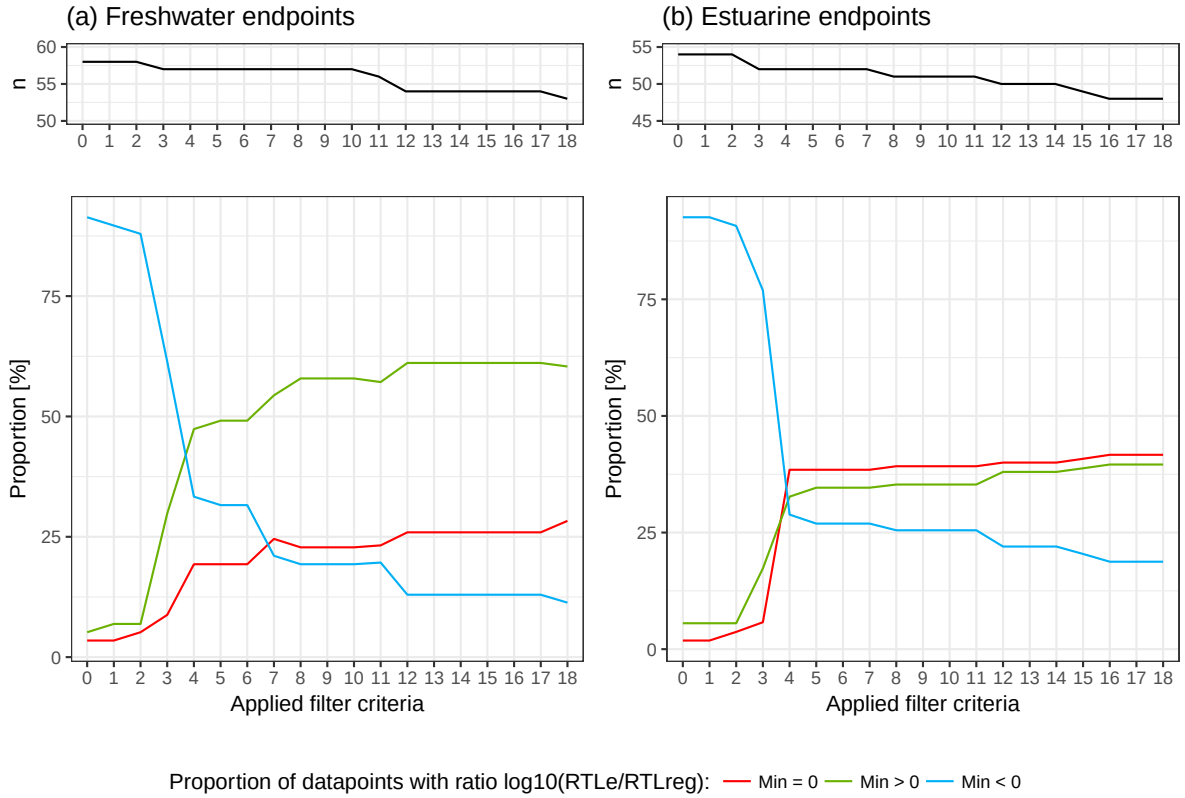


Figure 4: Influence of additive filter step application on the position of the most sensitive endpoints per substance expressed as proportion of endpoints with a ratio $\log_{10}(RTL_e/RTL_{reg})$ being higher, equal or lower than 0 for freshwater (a) and estuarine (b) endpoints. The proportion of substances with a most sensitive endpoint that is smaller than the RTL_{reg} ($\log_{10}(RTL_e/RTL_{reg}) < 0$) decreases the more filter steps are applied. In turn, the proportion of substances for which threshold estimations are correct ($\log_{10}(RTL_e/RTL_{reg}) = 0$) and the proportion of substances with threshold estimates higher than the RTL_{reg} ($\log_{10}(RTL_e/RTL_{reg}) > 0$) increase. The number of remaining substances after additive filter step applications is displayed in the upper panels. The more filter steps are applied, the more substances are excluded.

3.1.2 Threshold estimates from the null, mid and full model

After filter criteria were chosen, the second part of the model, namely the computation of threshold values from the pre-selected datapoints, was evaluated, this time on substance level. The full model, mid model and null model were applied to ECOTOX data filtering for the CAS numbers of substances included in the calibration dataset. For each substance, the position of the endpoint used for regulatory threshold value generation was compared to the remaining set of datapoints after filter applications (fig. 5-7). The more strict the filter criteria are selected, the less endpoints remain (orange endpoints) which lie below the endpoint used as regulatory threshold value (blue cross). The distributions of endpoints per substance yielded the median regulatory threshold positions with confidence intervals among all substances: For freshwater organisms, median endpoint positions expressed as quantiles lay at 0.09 (CI 0.07 - 0.16) in the null model. In the mid model and the full model, the estimation of endpoint positions within the distribution of all data points per substance decreased to 0.05 (CI 0.0 - 0.08) and 0.0 (CI 0.0 - 0.11), respectively. Since in the full model, the proportion of underpredictions was lowest and the proportion of estimates that met the RTL_{reg} was highest, if compared to the mid and the null model (fig. 4, 5-7), the most sensitive endpoint from full model filter criteria applications was chosen to be most suitable for threshold predictions. For estuarine models, the median position of regulatory endpoints in the null model lay at the 0.27 quantile (CI 0.15 - 0.33), in the mid model at the 0.20 quantile (CI 0.11 - 0.25), and in the full model at the 0.25 quantile (CI 0.0 - 0.33). Here, the median position of regulatory endpoints lay at the 0.33 quantile, but in a substantial proportion of substances, the position of the regulatory endpoint approached the 0 quantile (fig. 8), and the 0 quantile lay also at the lower end of the 95% confidence level. Thus, for the estuarine model, also the most sensitive endpoint was chosen for threshold predictions.

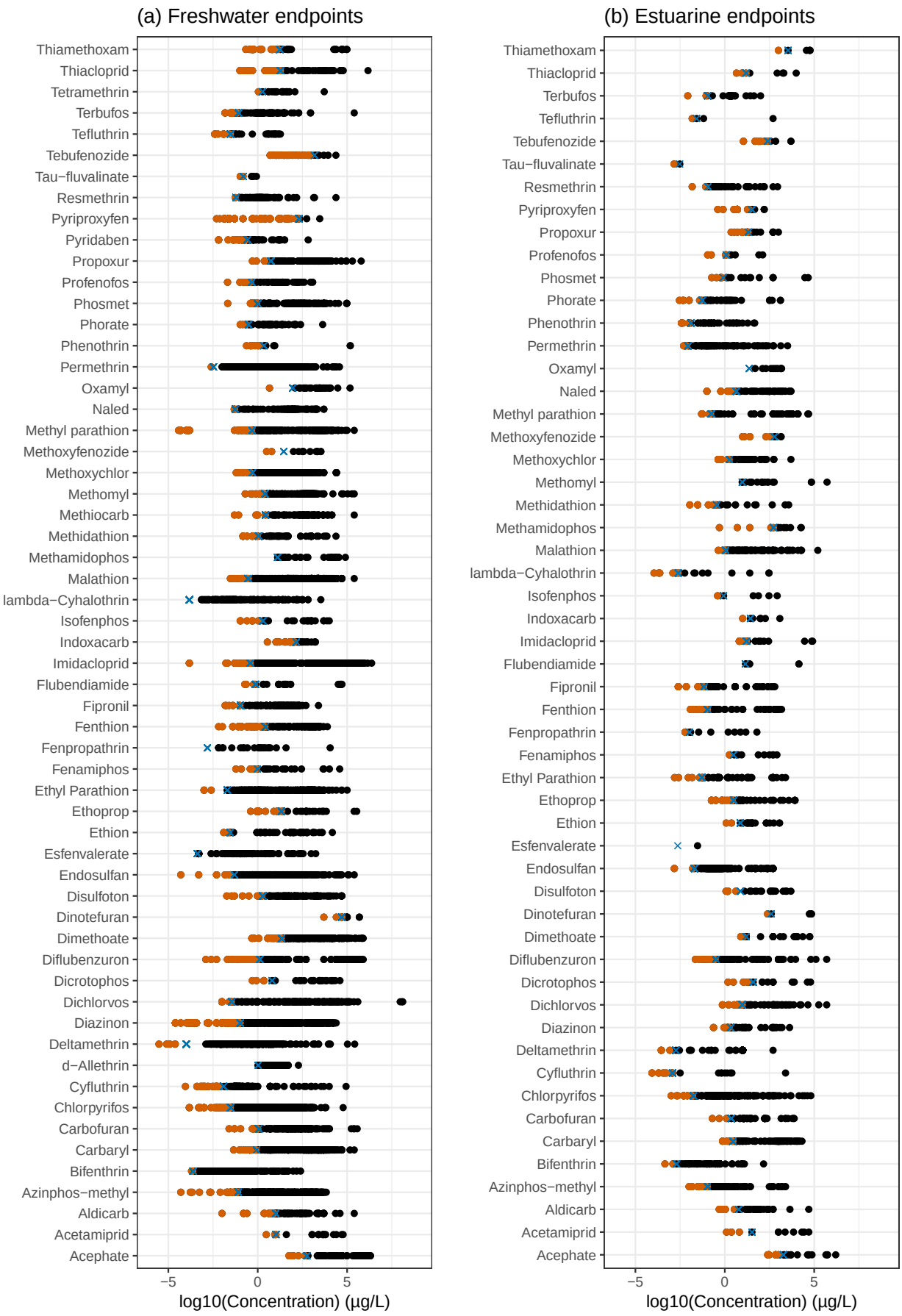


Figure 5: Endpoints remaining after null model filter applications for (a) freshwater and (b) estuarine endpoints per insecticide. Black dots indicate endpoints which would be higher than the relevant endpoint used for threshold generations (blue cross) and orange dots indicate endpoints which would be lower than the relevant endpoint used for threshold generations.

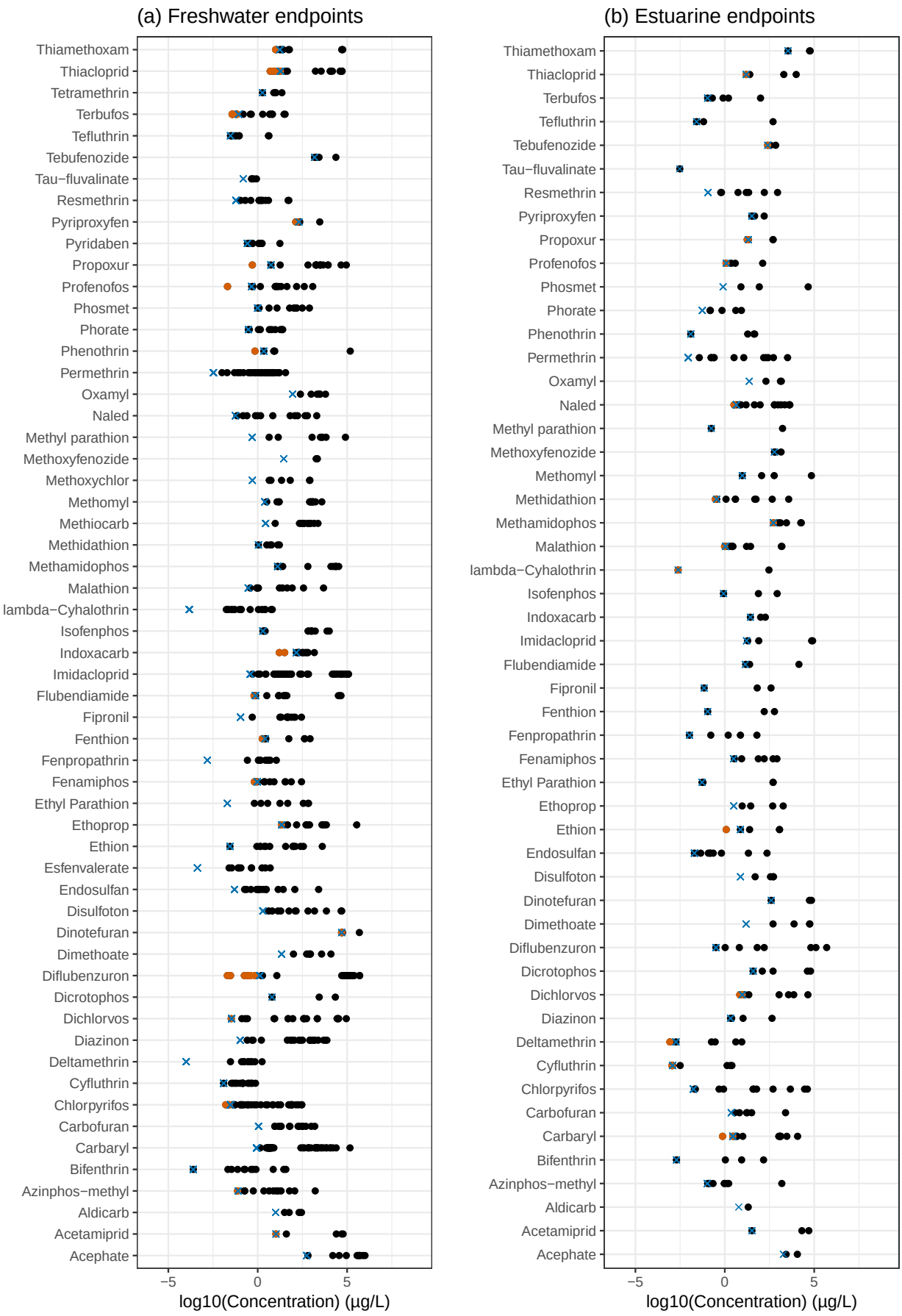


Figure 6: Endpoints remaining after mid model filter applications for (a) freshwater and (b) estuarine endpoints per insecticide. Black dots indicate endpoints which would be higher than the relevant endpoint used for threshold generations (blue cross) and orange dots indicate endpoints which would be lower than the relevant endpoint used for threshold generations.

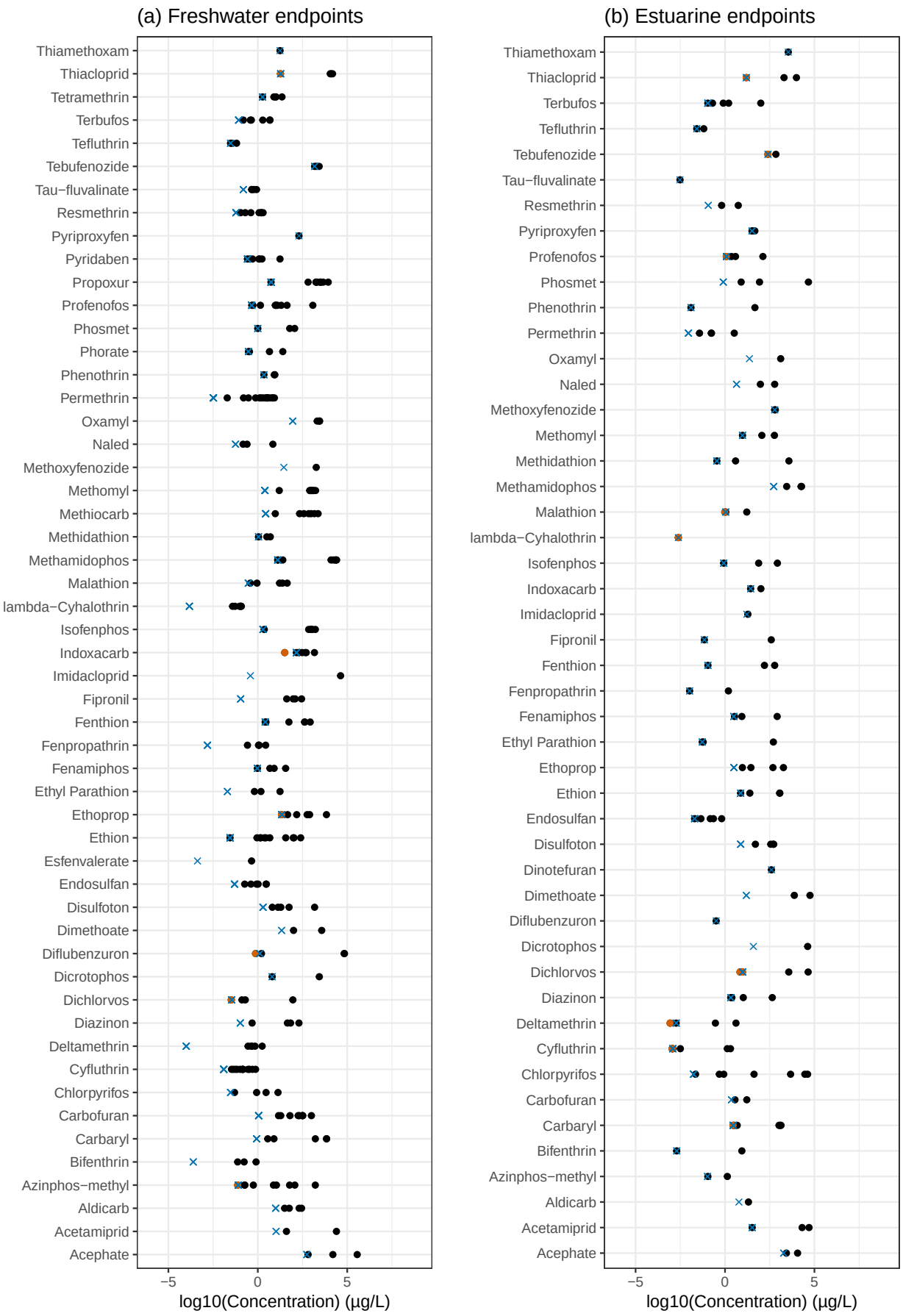
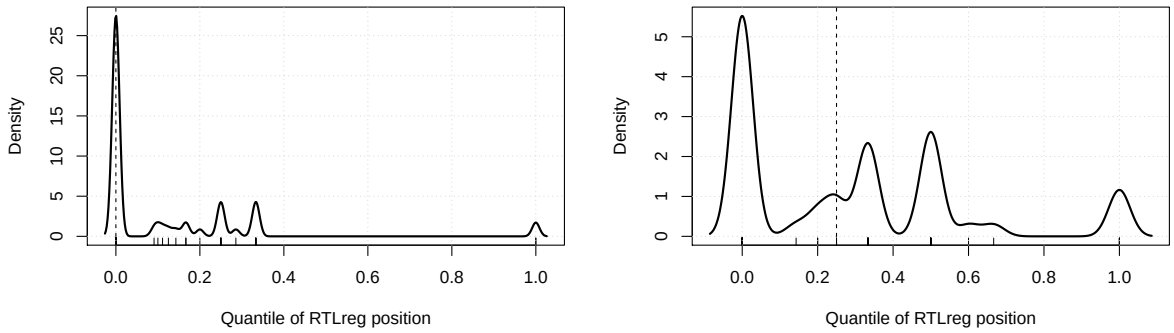


Figure 7: Endpoints remaining after full model filter applications for (a) freshwater and (b) estuarine endpoints per insecticide. Black dots indicate endpoints which would be higher than the relevant endpoint used for threshold generations (blue cross) and orange dots indicate endpoints which would be lower than the relevant endpoint used for threshold generations.



(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure 8: Distribution of RTL_{reg} position among insecticides for (a) freshwater endpoints and (b) estuarine endpoints with median (dashed lines) for full model filter applications. For freshwater organisms, median endpoint positions expressed as quantiles lay at 0.0 (CI 0.0 - 0.11). For estuarine models, the median quantile lay at 0.25 (CI 0.0 - 0.33) whereby a large proportion of RTL_{reg}s approach the lowest endpoint (0. quantile).

3.2 Model validation

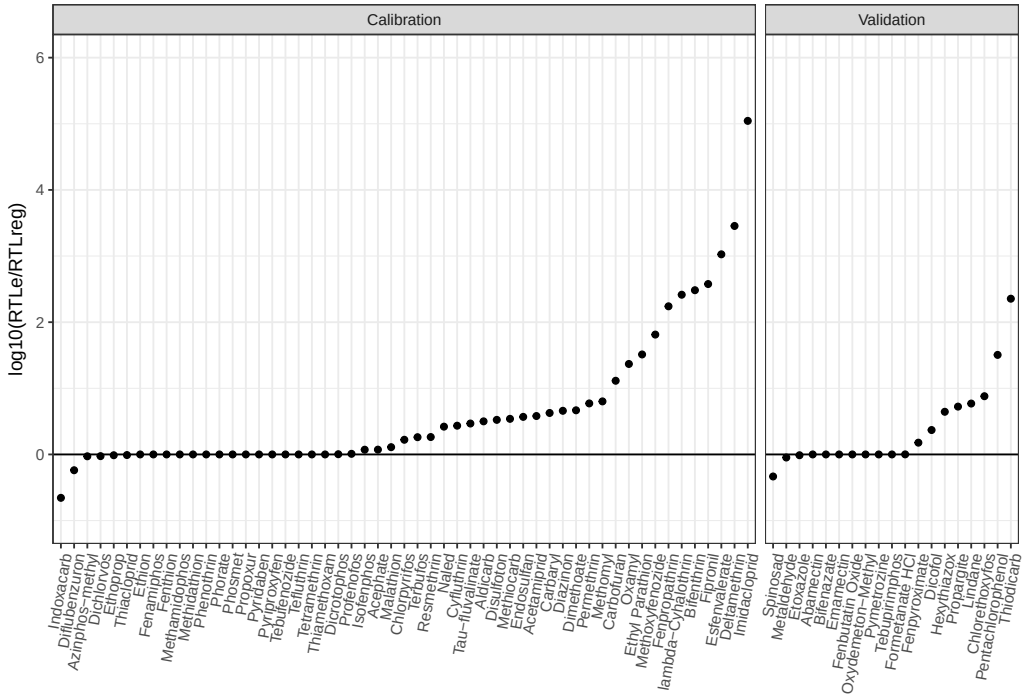
3.2.1 Model predictions

The precision of RTL_e predictions for calibration data and validation data was compared (fig. 9). Again, the ratio between RTL_e and RTL_{reg} was used as a measure to evaluate the goodness of model performance. While the number of insecticides for which RTL_e predictions underestimated the respective RTL_{reg} was low, a total of 13 and 5 RTL_es exceeded the respective RTL_{reg} by more than one order of magnitude in freshwater and estuarine model applications, respectively (fig.9), in both calibration and validation datasets. Even though the validation dataset was smaller than the calibration dataset, the distribution of datapoints appeared to be comparable, what could be confirmed by hypothesis testing. Since assumptions of standard students t-tests (normal distribution and variance homogeneity) were not met for the $\log_{10}$-transformed validation and calibration data ($\log_{10}(RTL_e/RTL_{reg})$), a nonparametric Wilcoxon Rank Sum test was conducted to analyze potential differences in predictions from the calibration and validation datasets. For freshwater and estuarine model predictions, hypothesis testing revealed no significant differences between predictions from the validation and calibration datasets (W = 428.5, *p*-value = 0.334, and W = 344, *p*-value = 0.895, respectively).

The comparison of included versus non-included datapoints in the database revealed a significant difference of prediction precision between substances for which endpoints were listed in the database and endpoints that were not included in the database for freshwater and estuarine models (W = 184, *p*-value < 0.001, W = 187, *p*-value = 0.03, respectively). Thus, if

datapoints were included in the database, predictions were significantly better than if the relevant endpoint was not included. This is in line with the distribution of data in figure 10, where the majority of predictions for data included in the database met the RTL_{reg}. Further, for freshwater endpoints, only about one third of endpoints of the available data (validation and calibration

(a) Freshwater endpoints



(b) Estuarine endpoints

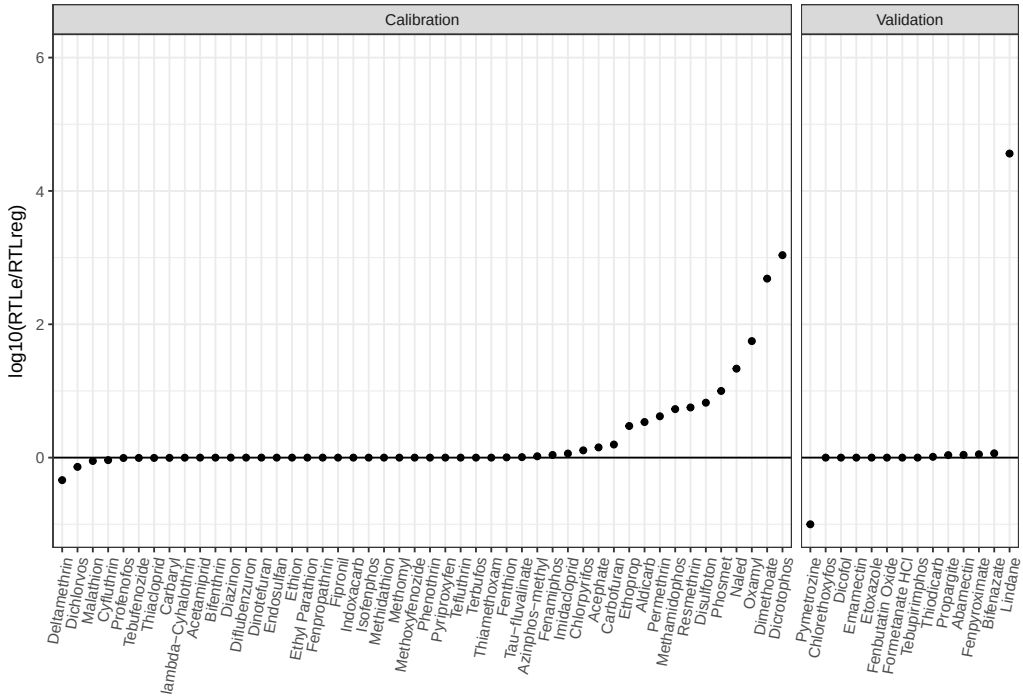
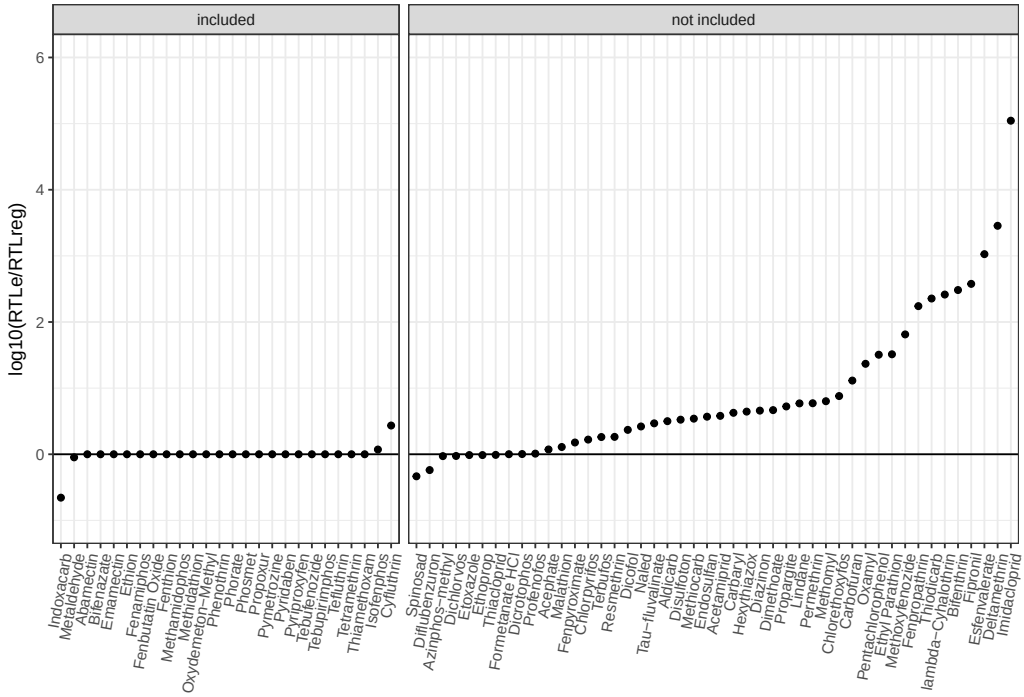


Figure 9: Evaluation of goodness of RTL_e predictions (expressed as $\log_{10}(RTL_e/\ RTL_{reg})$) for validation and calibration data for (a) freshwater and (b) estuarine full model predictions.

datasets) had actually endpoints equalling the relevant endpoint reported in the regulatory documents, whereas for estuarine model predictions, about half of the substances analyzed reported the correct endpoint in the database.

(a) Freshwater endpoints



(b) Estuarine endpoints

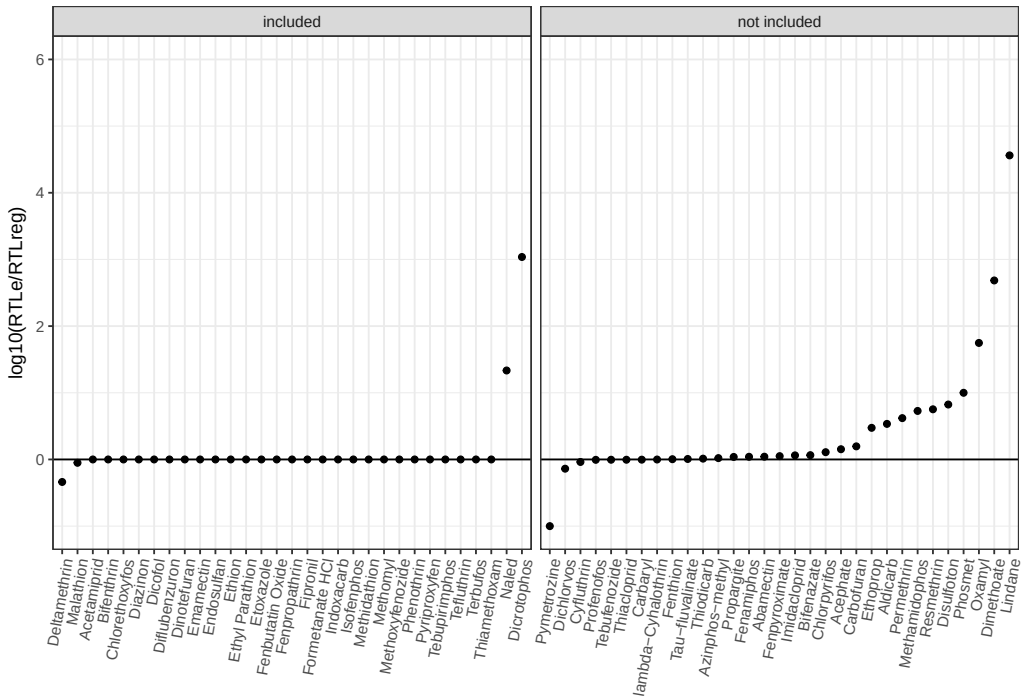


Figure 10: Evaluation of goodness of RTL_e predictions (expressed as $\log_{10}(RTL_e/RTL_{reg})$) for included versus non-included endpoints in the ECOTOX database, which that equal the RTL_{reg} if LOCs were applied, for (a) freshwater and (b) estuarine full model predictions.

3.2.2 Relationship between prediction precision and potential explanatory variables

Correlations between the single predictors were calculated to explore potential linear relationships. Results are summarized in table 2 and 3 for freshwater and estuarine model predictions, respectively. Pearson's correlation coefficients were computed once for all predictor variables against RTL_e, RTL_{reg} and the ratio $\log_{10}(RTL_e/RTL_{reg})$ as response variable. $Log_{10}(RTL_e)$ and $\log_{10}(RTL_{reg})$ were highly correlated (r = 0.83 and r = 0.90) in freshwater and estuarine model predictions, respectively, and correlations were significant (table 2 and 3). Testing the correlation between the $\log_{10}(RTL_{reg})$ and $\log_{10}(RTL_e/RTL_{reg})$ revealed a strong relationship for freshwater model applications (r = -0.45, p-value < 0.001) and no relationship for estuarine model applications (r = 0.01, p-value = 0.947). The degree of association between $\log_{10}(RTL_e)$ and $\log_{10}(RTL_e/RTL_{reg})$ was weak (r = 0.12), but non-significant for freshwater model applications, and strong (r = -0.44, p-value < 0.001) for estuarine model application. The RTL_{reg} and RTL_e as measures of toxicity were correlated with the majority of physical-chemical parameters. However, this was not evident for the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision in the estuarine model, and only for $\log_{10}(RTL_e/RTL_{reg})$ against the molecular weight in the freshwater model (fig. A3-5, appendix). The abundance of datapoints per substances was not associated with prediction precision for freshwater models, but for estuarine models, where Pearson's correlation coefficient revealed a weak significant association of the parameters (r = -0.26, p-value = 0.038; fig. A6, appendix). The year of publication was positively correlated with prediction precisions in freshwater models (r = 0.29, p-value = 0.014), but not in estuarine models (fig. A7, appendix).

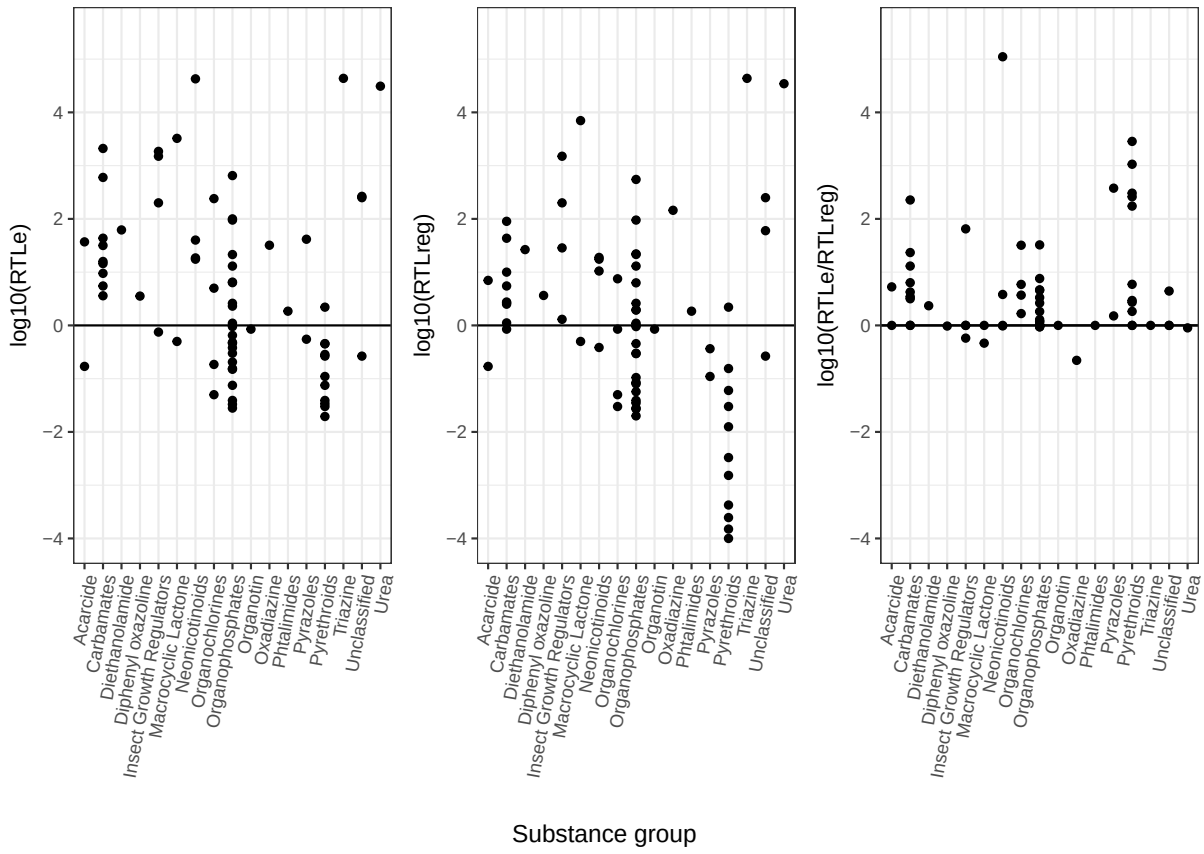
The graphical inspection of RTL_e, RTL_{reg} and the ratio $\log_{10}(RTL_e/RTL_{reg})$ against the categorical parameters *substance group* and *dataset type* revealed that patterns detected for RTL_e and RTL_{reg} did not translate into the measure of model prediction precisions ($\log_{10}(RTL_e/RTL_{reg})$, fig. 11 & 12). For instance, for the *substance groups*, the plots indicated that pyrethroids were especially toxic, whereas carbamates were less toxic, but there was no substance group of insecticides for which prediction precision was worse if compared to the other substance groups with regard to the number of predictions that underestimated the RTL_{reg}s. However, if RTL_es were plotted as a function of RTL_{reg}s, the group of pyrethroids appeared to be clustered in freshwater and estuarine model predictions, whereby RTL_{reg}s were not underestimated in the freshwater model application but rather predicted too high, being the most prominent substance group with insecticides that were overestimated by more than two orders of magnitude (fig. 9 and 11). The estuarine model application resulted only once in an underestimation of the RTL_{reg}, and most of the other pyrethroids were estimated correctly. All other substance groups were evenly scattered, such that no further clusters were evident (fig. A8-A10, appendix).

Table 2: Freshwater RTL_es predictions: Pearson's product-moment correlation for predictor variables against response variables (once for RTL_es and once for the ratio $\log_{10}(RTL_e/RTL_{reg})$). Bold values are statistically significant (p<0.05)

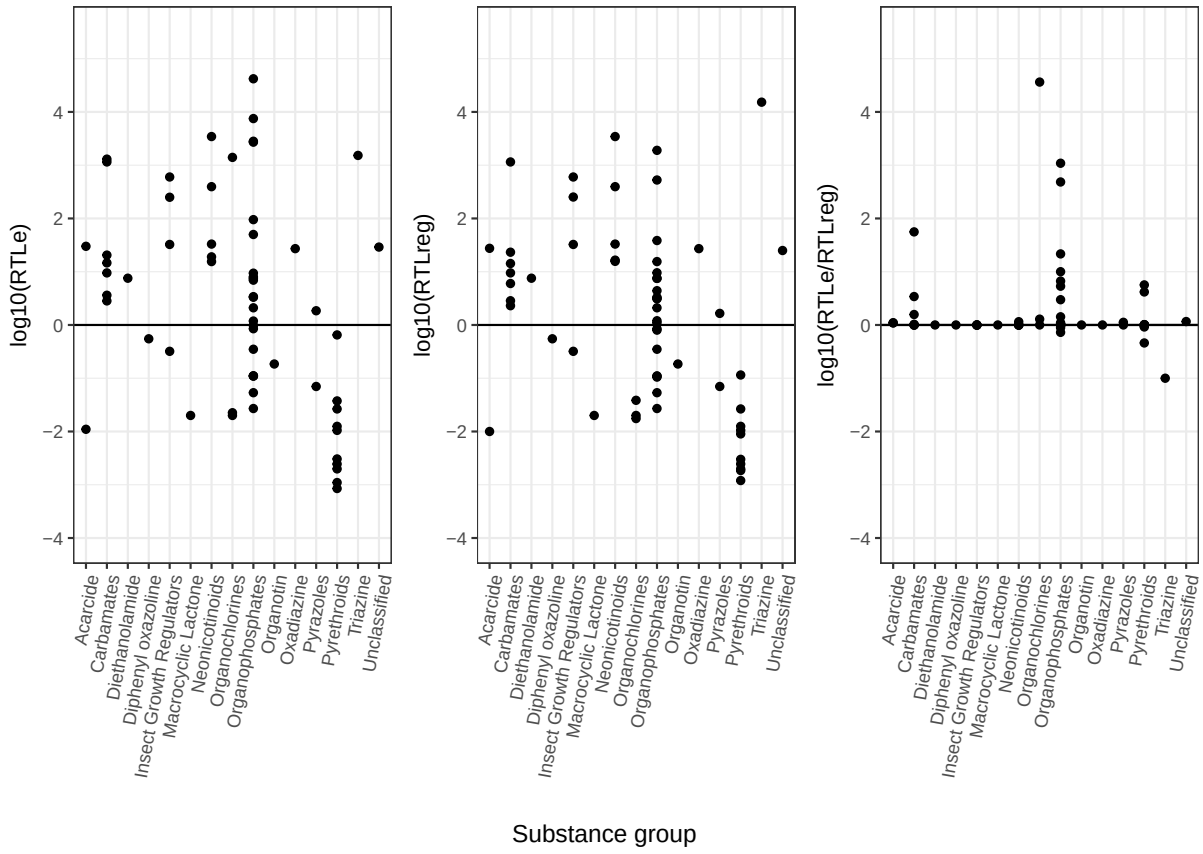
Freshwater Data	Pearson's *r* (95%-CI)	*p*-value	*df*	*t*
$\log_{10}(RTL_e)$: $\log_{10}(RTL_{reg})$	0.83 (0.74; 0.89)	**< 0.001**	70	12.43
$\log_{10}(RTL_e)$: $\log_{10}$(ratio)	0.12 (-0.11; 0.34)	0.303	70	1.04
$\log_{10}(RTL_e)$: MW	-0.37(-0.58; -0.11)	**0.006**	52	-2.86
$\log_{10}(RTL_e)$: K_{OC}	-0.26 (-0.56; 0.12)	0.174	28	-1.40
$\log_{10}(RTL_e)$: $\log_{10}$(Solubility)	0.35 (0.05; 0.59)	**0.022**	40	2.37
$\log_{10}(RTL_e)$: n	-0.46 (-0.63; -0.26)	**< 0.001**	70	-4.37
$\log_{10}(RTL_e)$: reference year	0.13 (-0.11; 0.35)	0.288	70	1.07
$\log_{10}(RTL_{reg})$: $\log_{10}$(ratio)	-0.45 (-0.62; -0.25)	**< 0.001**	70	-4.24
$\log_{10}(RTL_{reg})$: MW	-0.49 (-0.67; -0.25)	**< 0.001**	52	-4.02
$\log_{10}(RTL_{reg})$: K_{OC}	-0.43 (-0.68; -0.08)	**0.019**	28	-2.49
$\log_{10}(RTL_{reg})$: $\log_{10}$(Solubility)	0.45 (0.17; 0.66)	**0.003**	40	3.22
$\log_{10}(RTL_{reg})$: n	-0.33 (-0.52; -0.11)	**0.004**	70	-2.95
$\log_{10}(RTL_{reg})$: reference year	-0.05 (-0.28; 0.19)	0.693	70	-0.40
$\log_{10}$(ratio) : MW	0.30 (0.03; 0.52)	**0.003**	52	2.23
$\log_{10}$(ratio) : K_{OC}	0.18 (-0.19; 0.51)	0.332	28	0.99
$\log_{10}$(ratio) : $\log_{10}$(Solubility)	-0.25 (-0.52; 0.06)	0.107	40	-1.65
$\log_{10}$(ratio) : n	-0.15 (-0.37; 0.09)	0.210	70	-1.26
$\log_{10}$(ratio) : reference year	0.29 (0.06; 0.49)	**0.014**	70	2.51

Table 3: Estuarine RTL_es predictions: Pearson's product-moment correlation for predictor variables against response variables (once for RTL_es and once for the ratio $\log_{10}(RTL_e/RTL_{reg})$). Bold values are statistically significant (p<0.05)

Estuarine Data	Pearson's *r* (95%-CI)	*p*-value	*df*	*t*
$\log_{10}(RTL_e)$: $\log_{10}(RTL_{reg})$	0.90 (0.84; 0.93)	**< 0.001**	60	16.07
$\log_{10}(RTL_e)$: $\log_{10}$(ratio)	0.44 (0.22; 0.62)	**< 0.001**	60	3.82
$\log_{10}(RTL_e)$: MW	-0.58 (-0.74; -0.36)	**< 0.001**	47	-4.93
$\log_{10}(RTL_e)$: K_{OC}	-0.64 (-0.82; -0.34)	**< 0.001**	25	-4.17
$\log_{10}(RTL_e)$: $\log_{10}$(Solubility)	0.53 (0.25; 0.73)	**< 0.001**	36	3.73
$\log_{10}(RTL_e)$: n	-0.34 (-0.54; -0.10)	**0.007**	60	-2.78
$\log_{10}(RTL_e)$: reference year	-0.03 (-0.28; 0.22)	0.798	60	-0.26
$\log_{10}(RTL_{reg})$: $\log_{10}$(ratio)	0.01 (-0.24; 0.25)	0.947	60	0.07
$\log_{10}(RTL_{reg})$: MW	-0.56 (-0.72; -0.33)	**< 0.001**	47	-4.60
$\log_{10}(RTL_{reg})$: K_{OC}	-0.58 (-0.79; -0.25)	**0.002**	25	-3.55
$\log_{10}(RTL_{reg})$: $\log_{10}$(Solubility)	0.53 (0.26; 0.73)	**< 0.001**	36	3.77
$\log_{10}(RTL_{reg})$: n	-0.25 (-0.47; 0.002)	0.052	60	-1.98
$\log_{10}(RTL_{reg})$: reference year	-0.02 (-0.27; 0.23)	0.849	60	-0.19
$\log_{10}$(ratio) : MW	-0.17 (-0.43; 0.12)	0.250	47	-1.16
$\log_{10}$(ratio) : K_{OC}	-0.35 (-0.64; 0.04)	0.075	25	-1.86
$\log_{10}$(ratio) : $\log_{10}$(Solubility)	0.05 (-0.27; 0.36)	0.772	36	0.29
$\log_{10}$(ratio) : n	-0.26 (-0.48; -0.02)	**0.038**	60	-2.12
$\log_{10}$(ratio) : reference year	0.03 (-0.27; 0.23)	0.843	60	-0.20

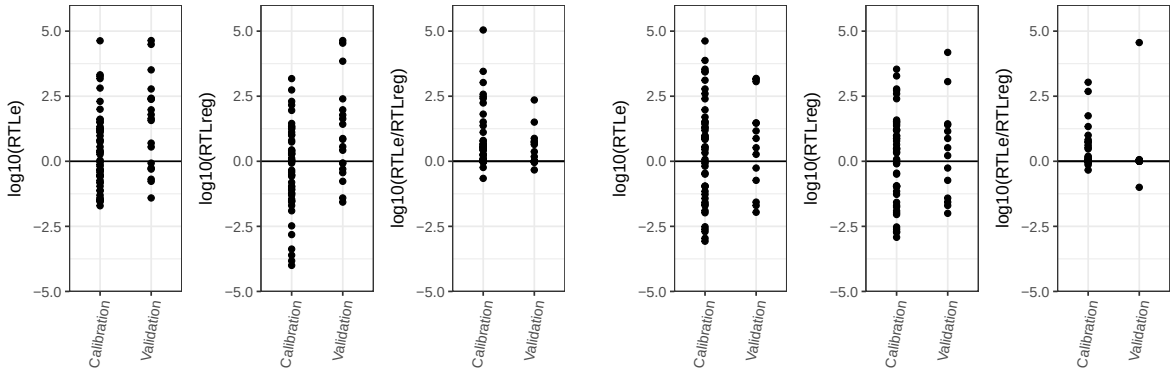


(a) Freshwater endpoints



(b) Estuarine endpoints

Figure 11: Substance group versus RTL_e, RTL_{reg} and the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision for freshwater (a) and estuarine endpoints (b).

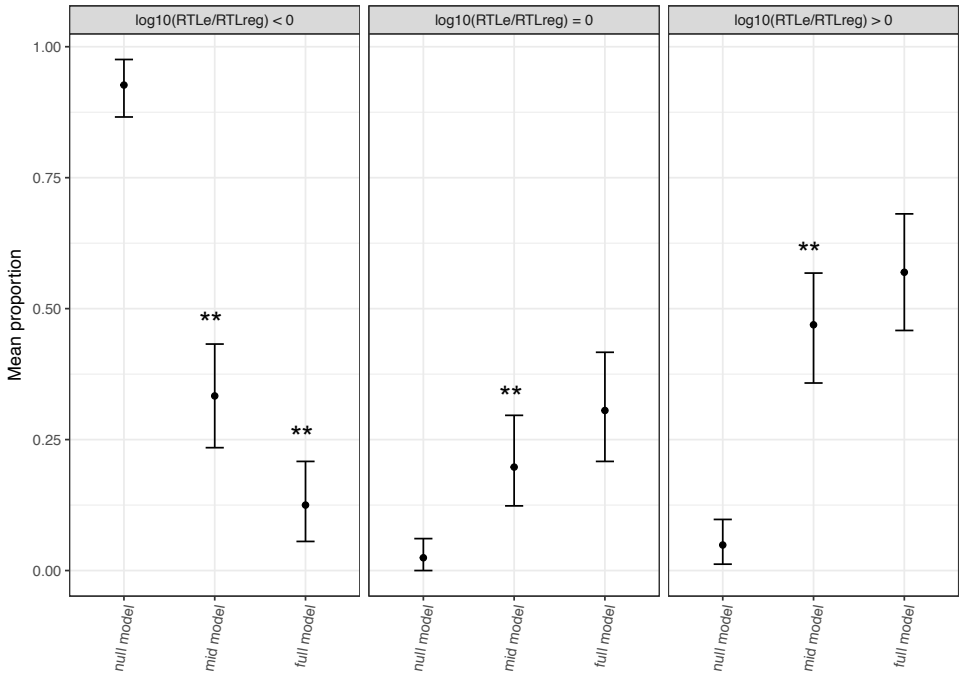


(a) Freshwater endpoints **(b)** Estuarine endpoints

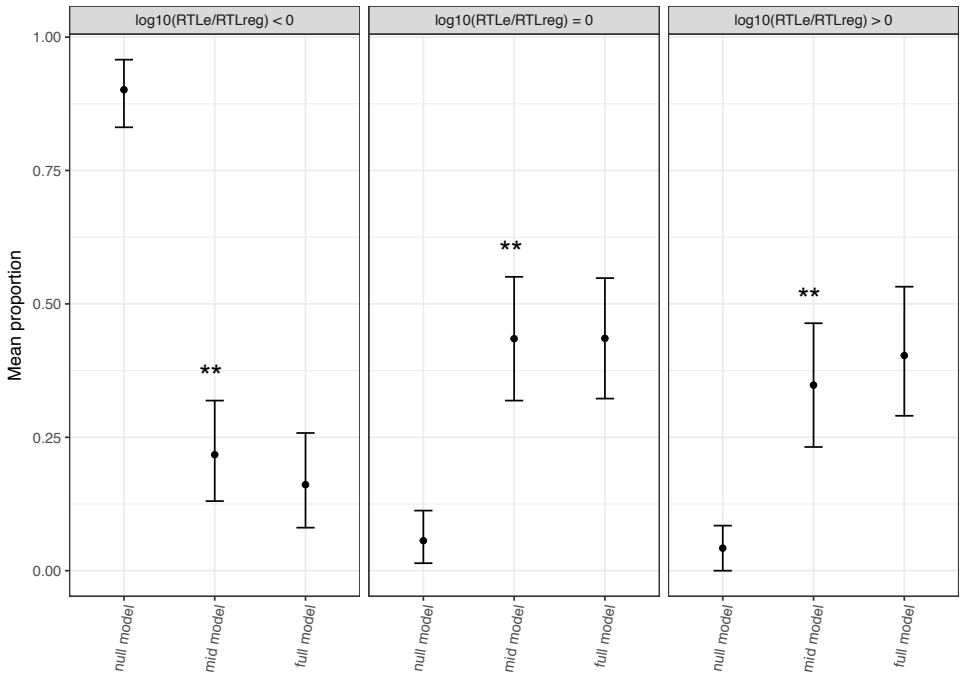
Figure 12: Priority group versus RTL_e, RTL_{reg} and the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision for freshwater (a) and estuarine endpoints (b).

3.2.3 Evaluation of model precision

To evaluate the model precision, bootstrapping provided mean proportions of RTL_e predictions which over- or underestimated the RTL_{reg}s and the mean proportion of estimated endpoints (RTL_es) which met the actual regulatory endpoint with 95% confidence intervals (table A1a, appendix). Further, the influence of whether endpoints are listed in the ECOTOX database (inclusion versus non-inclusion) on model precision was evaluated (table A1b, appendix). If the null model was applied to estimate threshold levels for a set of substances, the majority of endpoints (92.7% (CI 86.6-97.6) and 90.1% (CI 83.1-90.1) for freshwater and estuarine models, respectively) would be estimated too low. In the mid model, prediction precision was higher, as a significantly lower proportion underestimated the RTL_{reg}, a significantly higher proportion met the RTL_{reg}, although a significantly higher proportion exceeded the RTL_{reg} (p<0.01, estimated by non-overlapping confidence levels, fig. 13). From the mid model to the full model, predictions improved again, even though only the proportion of underpredictions decreased significantly (p<0.01, estimated by non-overlapping confidence levels, fig. 13). While the presence of datapoints in the ECOTOX database improved predictions significantly (p<0.01, estimated by non-overlapping confidence levels, higher proportion of predictions with a $\log_{10}(RTL_e/RTL_{reg}) = 0$ and $\log_{10}(RTL_e/RTL_{reg}) > 0$ if compared to the full model), the proportion of predictions that underestimated the RTL_{reg} (with $\log_{10}(RTL_e/RTL_{reg}) < 0$), was neither significantly different from the proportion of substances, that were not included in the database, nor from predictions of the full model, that comprised predictions for substances with included and non-included actual endpoints (significance inferred from confidence intervals, fig.14).

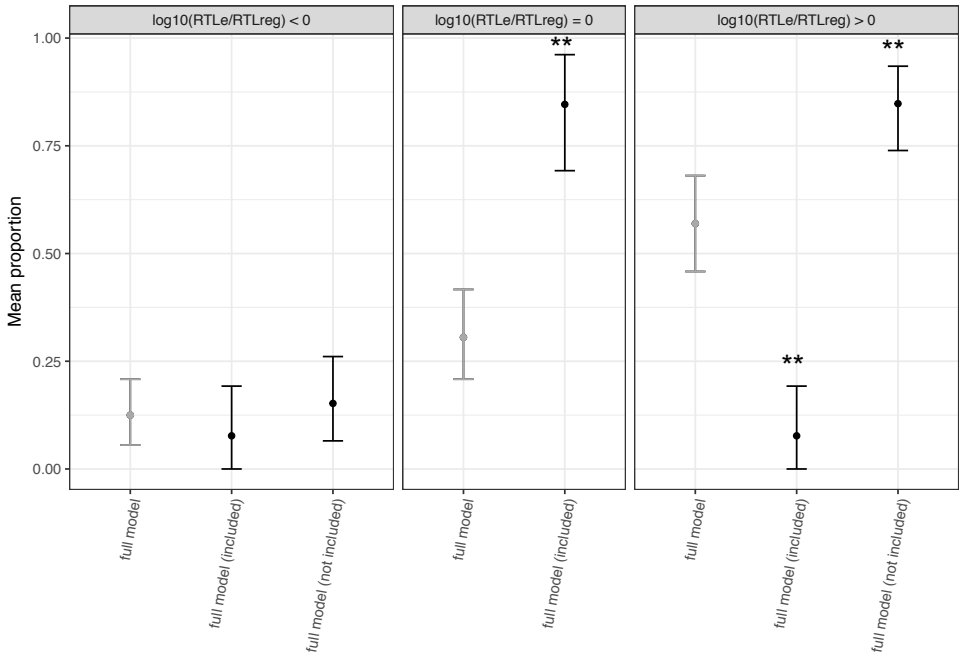


(a) Freshwater endpoints

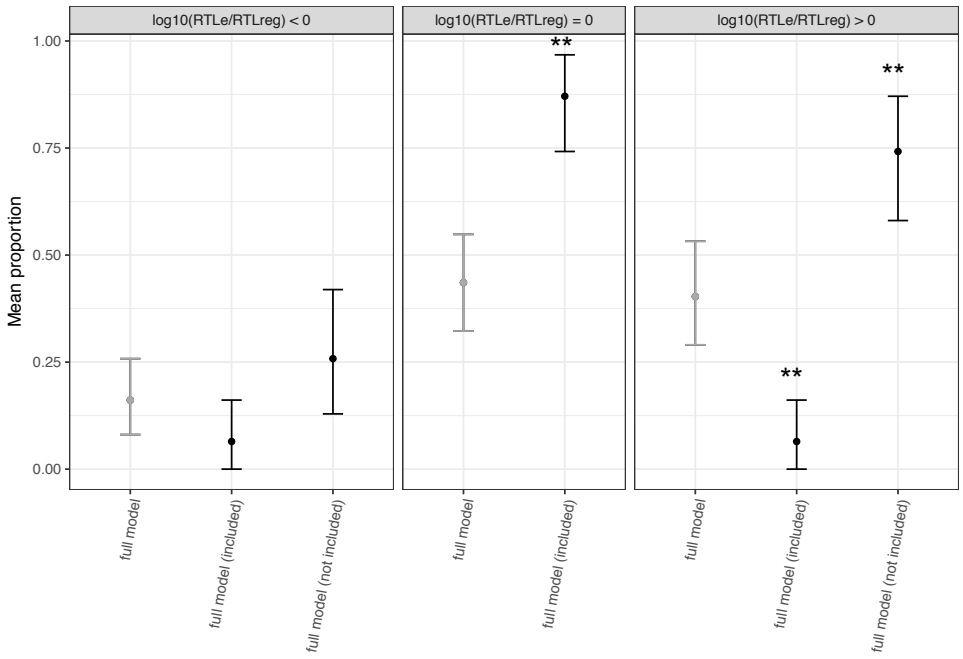


(b) Estuarine endpoints

Figure 13: Mean prediction precision for null models, mid models, and full models for freshwater endpoints (a) and estuarine endpoints (b). Points indicate the mean proportion of RTL_es, which underpredict ($\log_{10}(RTL_e/RTL_{reg}) > 0$), meet ($\log_{10}(RTL_e/RTL_{reg}) = 0$) or overpredicted ($\log_{10}(RTL_e/RTL_{reg}) > 0$) the actual RTL_{reg}s for a set of substances. Whiskers indicate 95% bootstrapped confidence levels. Asterisks show significant differences (***** if p-value < 0.05, ****** if p-value < 0.01, Cumming & Finch, 2005) with regards to the respective previous model (null to mid model; mid to full model).



(a) Freshwater endpoints



(b) Estuarine endpoints

Figure 14: Difference between model prediction precision for listed and non-listed endpoints in the ECOTOX database for full model applications for freshwater endpoints (a) and estuarine endpoints (b). Points indicate the mean proportion of RTL_es, which underpredict ($\log_{10}(RTL_e/RTL_{reg}) > 0$), meet ($\log_{10}(RTL_e/RTL_{reg}) = 0$) or overpredicted ($\log_{10}(RTL_e/RTL_{reg}) > 0$) the actual RTL_{reg}s for a set of substances. Whiskers indicate 95% bootstrapped confidence levels. Asterisks show significant differences (***** if p-value < 0.05, ****** if p-value < 0.01, Cumming & Finch, 2005) with regards to the full model.

3.3 Model application

3.3.1 Estimations for a set of insecticides

The model application yielded five freshwater RTL_es and five estuarine RTL_es (table 4). The query with 23 further insecticides did not yield threshold estimates. For a full list of substances for which the model was applied (n = 29), please refer to table A2, in the appendix.

Table 4: Results of model application: RTL_es for set of insecticides, for which no regulatory documents were available

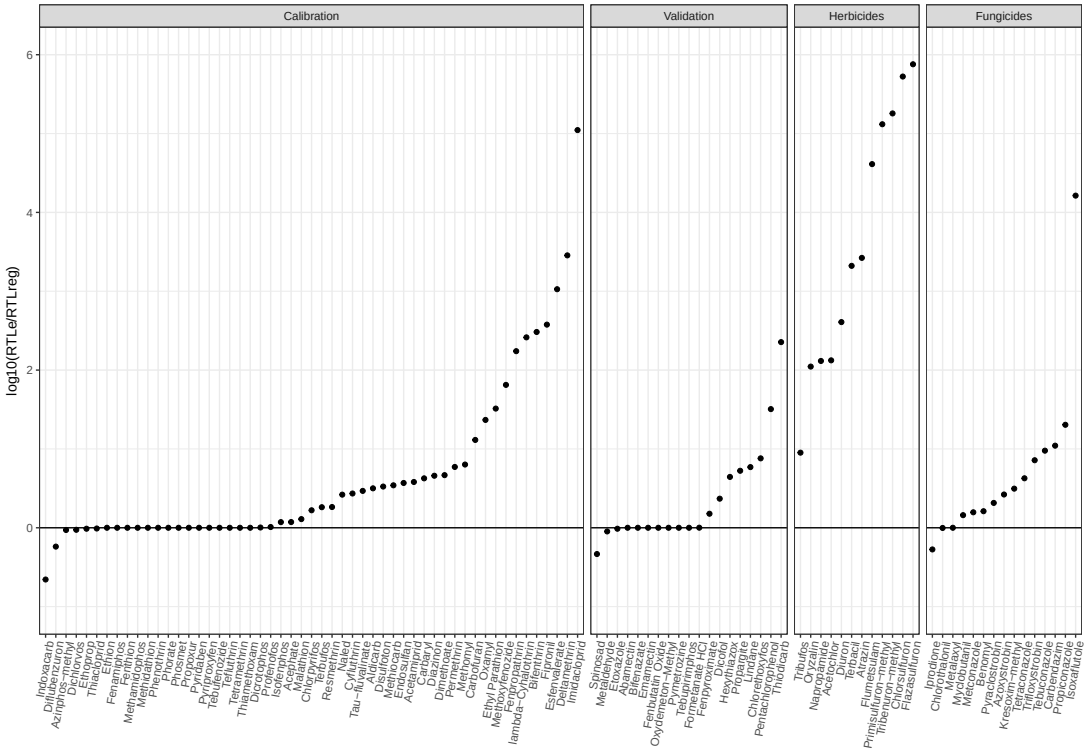
Freshwater predictions:			**Estuarine predictions:**		
Substance	**CAS number**	**RTL_e (ug/L)**	**Substance**	**CAS**	**RTL_e (ug/L)**
Fonofos	944229	**1.00**	Fonofos	944229	**0.25**
Mevinphos	7786347	**1.75**	Mevinphos	7786347	**0.65**
Bendiocarb	22781233	**14.60**	Bendiocarb	22781233	**3.35**
Fenvalerate	51630581	**0.03**	Fenvalerate	51630581	**0.70**
Demeton	8065483	**13.50**	Novaluron	116714466	**0.07**

3.3.2 Model extension for herbicides and fungicides

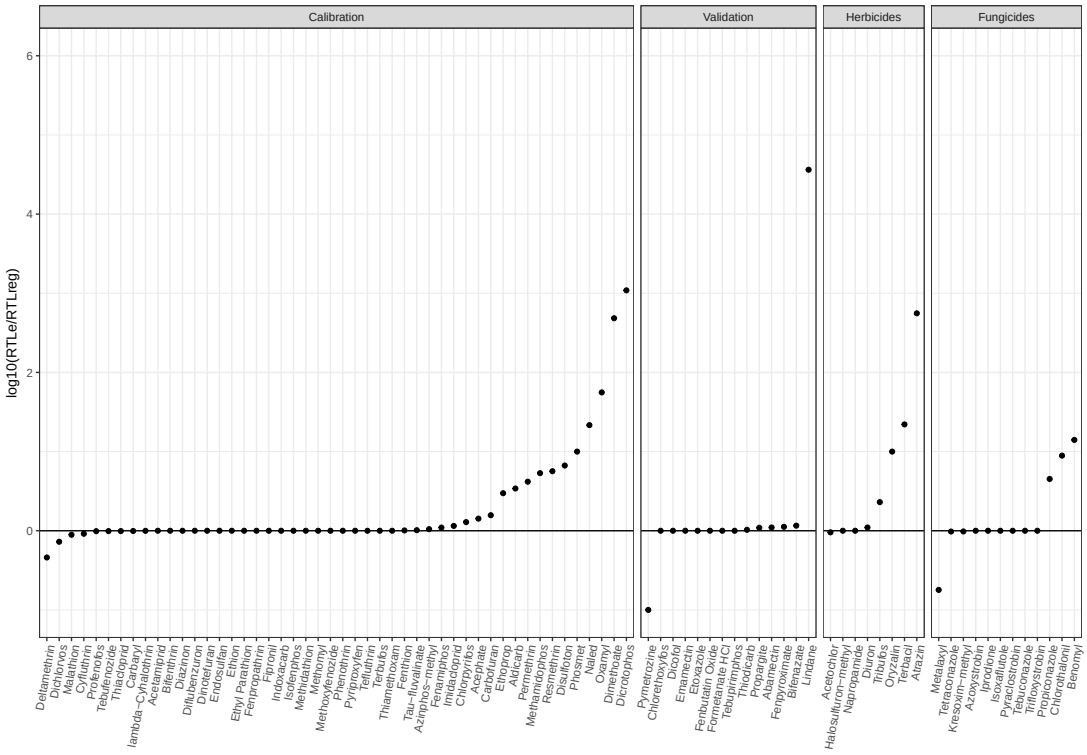
The original full model was adapted to allow aquatic plant species, effect types and study durations which are typical for aquatic plant studies (table 1, extended model filter criteria). The performance of model predictions for herbicides was significantly improved, if plant endpoints were considered (comparison of original herbicide predictions versus adapted herbicide predictions: W = 224, p-value < 0.001). Through the inclusion of plant endpoints into the model, the difference in model precision between insecticide predictions and herbicide predictions in the original full model (comparison of original insecticide predictions versus original herbicide predictions: W = 45, p-value < 0.001) disappeared (comparison of original insecticide predictions versus adapted herbicide predictions: W = 771, p-value = 0.626, fig. 15 (a) and 16 (a)). For fungicides, inclusion or exclusion of plant endpoints did not result in significant changes of prediction precision (comparison of original fungicide predictions versus adapted fungicide predictions: W = 149.5, p-value = 0.127), even though predictions were slightly better, if plant endpoints were considered in the model (fig. 16 (a)). Fungicide predictions did neither differ significantly from insecticide prediction precision if aquatic plants were considered (comparison of original insecticide predictions versus adapted fungicide predictions: W = 632, p-value = 0.296) or not (comparison of original insecticide predictions versus original fungicide predictions: W = 449, p-value = 0.305). Further, the inclusion of aquatic plant endpoints for herbicide predictions enabled the analysis of more substances if compared to the original full model predictions, since the amount of available substances increased from 12 to 20. For fungicides, the number of substances did not differ if aquatic plant endpoints were considered or not, such that in both

cases, predictions for 15 fungicides could be obtained. Full model adaptions are presented in the appendix A2.5.

For estuarine endpoints, predictions did not differ if aquatic plant endpoints were considered or if not (comparison of original model predictions and adapted model predictions for herbicides: W = 35, *p*-value = 0.7908; and fungicides: W = 72.5, *p*-value = 1, fig. 16 (b)). Further, the herbicide and fungicide prediction precision did not differ significantly from the insecticide predictions if aquatic plant endpoints were not considered (comparison of original insecticide predictions versus original herbicide predictions: W = 175.5, *p*-value = 0.168; and comparison of original insecticide predictions versus original fungicide predictions: W = 426, *p*-value = 0.411, fig. 15 (b) and 16 (b)), and if plant endpoints were considered for model predictions (comparison of original insecticide predictions versus adapted herbicide predictions: W = 187.5, *p*-value = 0.251; and comparison of original insecticide predictions versus adapted fungicide predictions: W = 426, *p*-value = 0.411). For estuarine model predictions, the amount of different substances for which predictions are possible does not depend on the inclusion or exclusion of aquatic plant endpoints. For herbicides, in both cases, 8 substances are available. For fungicides, in both cases, there is data for 12 different substances. Since estuarine model adaptions did not result in a substantial improvement of the model, the original estuarine model is kept for fungicide and herbicide threshold predictions.



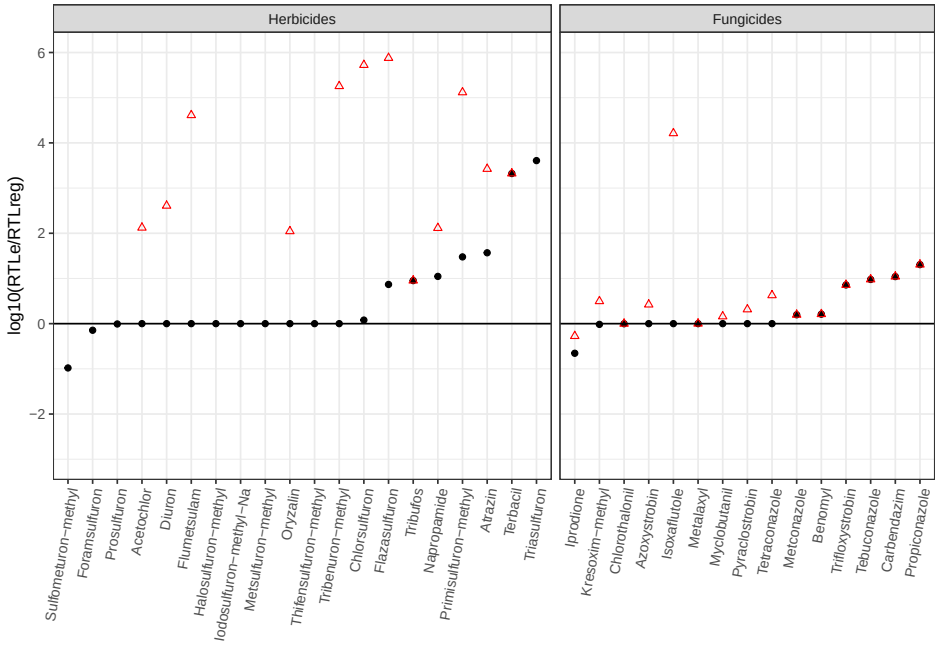
(a) Freshwater endpoints



(b) Estuarine endpoints

Figure 15: Comparison of model prediction precision for validation, calibration, and the extension data sets (herbicide and fungicide data) expressed as the ratio $\log_{10}(RTL_e/RTL_{reg})$ for freshwater (a) and estuarine (b) endpoints.

(a) Freshwater endpoints



(b) Estuarine endpoints

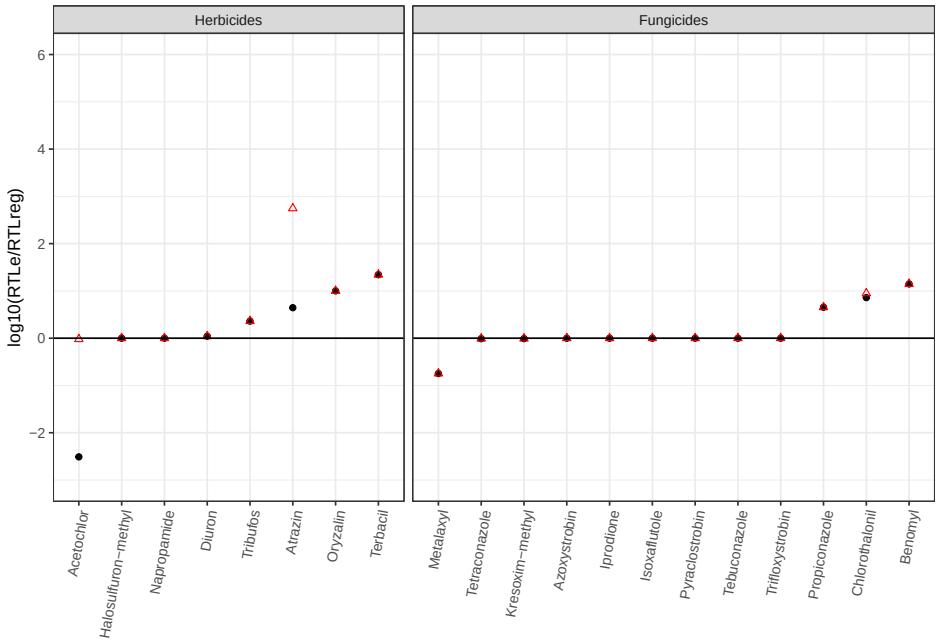


Figure 16: Comparison of model prediction precision of the extension data sets (herbicide and fungicide data) expressed as the ratio $\log_{10}(RTL_e/RTL_{reg})$ for freshwater (a) and estuarine (b) endpoints if aquatic plant endpoints were excluded (red triangles) or if aquatic plant filter criteria were included to the adapted full model (black dots) (for filter criteria, please refer to table 1, *Aq. plant 1-4*). Inclusion of aquatic plant filter criteria improved the performance of the freshwater model for herbicides and fungicides, but improvements were only significant for herbicide predictions (nonparametric Wilcoxon Rank Sum test with herbicide data: W = 224, p-value < 0.001; with fungicide data: W = 149.5, p-value = 0.127). In the adapted estuarine model, no difference was indicated if compared to the original full model (nonparametric Wilcoxon Rank Sum test with herbicide data: W = 35, p-value = 0.791; with fungicide data: W = 72.5, p-value = 1).

4 Discussion

4.1 Filter criteria selection

There were two different types of filters within the formulated 19 (0-18) filter criteria: the first group of filters helped to select only required study types (i.e, acute studies for freshwater and estuarine organisms exposed to insecticides and reported in an appropriate unit (null model and mid model, filters 0-5)), the other group comprised filters of technical nature and could be seen as the actual validity check of the pre-selected studies listed in the ECOTOX database. Thereby, the selection of the mainly technical validity criteria followed the principles given in the OCSPP standard test guidelines (US EPA, 2012*a*,*b*,*c*, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*) and the EPA guideline for review of the open literature (Brady, 2011), and their suitability and usefulness will be examined in more detail in the following.

Single endpoint, specification of applied method. For instance, filter 6 and 15 were formulated in accordance with the open literature guidelines (Brady, 2011), as it is mandatory for studies to be used in risk assessments to report a single endpoint, and to specify the applied test methods (e.g., as *U.S. EPA, OECD, GLP,* or *Other Method*, see ECOTOX User Guide, p. 56, (US EPA, 2017*c*), and ECOTOX Code Appendix, p. 695, (US EPA, 2017*b*)). Especially filter 15, which filters for studies with a specified test method, excludes a large proportion of datapoints, as 83 and 76% of endpoints are excluded from single filter application with regards to the null model, for freshwater and estuarine models, respectively. However, if filters were applied additively, the exclusion of studies without a specified test method (filter 15) does not lead to large changes in the remaining datapoints, such that most irrelevant endpoints excluded in the singular filter application are likely to be already excluded by the previous filter steps. If for measurements the code *MULT* is given, this means that multiple effects are reported as one result. According to ECOTOX Code Appendix (p. 606, (US EPA, 2017*b*)), if reviewers cannot distinguish separate endpoints, they use the code *MULT* to report the applied measurements. Therefore, it was concluded that these studies are not likely to be relevant for risk assessors, since no single endpoint is reported, that could be used for quantitative risk evaluation. In the RTLE model, this exclusion rational is implemented in filter 6, which is deemed relevant even though its application did not exclude a huge proportion of endpoints in the single filter application analysis, but increased the prediction precision in the freshwater model (fig 4., section 3.1.1).

Exclude comments. The effect comment criteria (filter 7) aimed at excluding study performance validity elements, as increased mortality in the controls, reported aeration, or further comments which would question the reliability of a study. According to the EFED guidance for

encoding toxicity data (US EPA, 2016*i*), in this field, effects other than the actual toxicity effect (e.g., related to pH, temperature or salinity) shall be coded. However, it appears as if this field is used to also refer to mortalities, report the measurement of mixture toxicity, or the aeration of test vessels. These criteria are prominent in study validity guidelines and the rejection rate analysis (US EPA, 1994; please also refer to tables A6-A9 in the appendix) as reasons for invalidation of studies, but hard to judge by applying filters to the encoded ECOTOX test properties only, since no standardized codes were used. Instead, it appears as if variable free text was entered here. Therefore, the presence of comments in this section was deemed adequate to exclude these studies, since comments most likely indicate a decrease in validity of the studies, such that it becomes less likely that studies were used for risk assessments. The relevance of this filter criteria was supported by the power analysis of single and additive filter application as well as precision of predictions, as the single application of filter 7 excluded 17 and 14% of endpoints with regard to the null model, for freshwater and estuarine models, respectively. In addition, the additive filter application indicated a slight improvement of remaining endpoints between filter step 7 and 8, and the prediction precision among all substances also slightly increased (Fig.4, proportions with a minimum below a $\log_{10}(RTL_e/RTL_{reg}) < 0$ decreased, while the proportion of $\log_{10}(RTL_e/RTL_{reg}) > 0$ slightly increased) if effect comments were not permitted. Unfortunately, it was not possible to formulate more precise filter criteria to exclude validity elements with regards to the test performance, what will be discussed in more detail subsequent to the current evaluation of filter criteria selection (section 4.1, see subsection *Difficulties with filter formulation*).

Test substance purity. Filter 8 and 9 were formulated to allow only studies which were conducted with a certain percentage of an active ingredient and excluding studies with a concentration type indicating the use of the formulated product. Both filter criteria improved the model performance as they excluded about 14 and 26% of endpoints in the freshwater model, and 11 and 17% of irrelevant endpoints in the estuarine model with regards to the null model, respectively. The rational behind this filter criteria was to exclude endpoints based on the product. The ECOTOX DB contains studies with toxicity data for the typical end-use product (TEP)/formulation or the active ingredient/technical grade active ingredient (TGAI), which might differ in their constitution and their toxicity. The product or formulation contains additional substances aside the actual active ingredient and might even contain further active ingredients. The OPP requires effect studies for the TGAI, and in some cases also for the TEP (US EPA,2015*b*; CFR40, §158.630), for instance if TEP toxicity is expected to exceed TGAI toxicity or if the end-use product is applied directly into aquatic environments (e.g., as some herbicides are). A higher toxicity of the product was only indicated for two of the about 60 reviewed in-

secticides (flubendiamide and resmethrin), for which tests with the TEP might have been more relevant than studies performed with the TGAI. However, the use of TGAI studies instead of TEP studies is not likely to decrease the reliability of the RTLE model predictions: if the TEP is more toxic than the TGAI (i.e., lower endpoints), but estimates are performed for the TGAI, the RTL_e is rather estimated too high (e.g., resmethrin, RTL_e/RTL_{reg}s = 1.83. For flubendiamide, in the full model no studies passed the validity-check, such that no estimations were possible.). Thus, it is not likely that the RTLE model overstates the risk by predicting too low threshold values. Therefore, it is deemed justifiable that the model only considers studies performed with the original active ingredient and requires a high purity of the test substance, whereby the 70% boundary was chosen iteratively from 100% downwards to obtain the best result while at the same time not losing too many potentially relevant endpoints.

Organisms source. Filter 10 and 13 were relevant to examine the source of organisms used for testing. Here, for reasons of test reliability, it is important to exclude test organisms with a potential pre-exposure in their life-history. Therefore, test sources which were not reported (*NR*) or given as *wild* were omitted. Further, since the aim was to filter results from acute standard tests, field studies were rejected, as the wanted acute standard test results mainly origin from highly standardized laboratory studies under controlled conditions (OCSPP standard guidelines). For fish, also wild caughts could be used, but only after acclimatization and from non-polluted sites (US EPA, 2016*g*). However, a missing or wrong acclimatization of test organisms was listed among the most important and most frequent reasons for study invalidations (table A9, 14a-b; (US EPA, 1994). Since the acclimatization periods could not be tracked easily within the provided ECOTOX code, to be conservative, wild test organism origins were not allowed in the present model. This is also supported by the single filter application analysis, given that filter 13 (omission of studies with test organisms of unknown sources) excluded 3 and 10% of datapoints with regard to the null model, for freshwater and estuarine models, respectively. Filter 10 (omission of studies other than lab studies) was less powerful, but still excluded 3 and 1% of datapoints with regard to the null model, for freshwater and estuarine models, respectively. From the additive filter application, no obvious improvements were apparent, even though the application of filter 10 removes one substance from the remaining set of data (fig. 4(a)), which had previously overestimated the actual RTL_{reg}. This leads to slight changes in the proportions of RTL_es meeting the RTL_{reg}, and RTL_es underestimating the RTL_{reg}, which are however solely based on the removal of one substance from the set of data. In conclusion, it appears to be reasonable to rather keep these filter criteria, even though only small improvements are obtained.

Time of publication. Filter 11 was included for validation and calibration purposes and re-

moved from the main model. Here, the publication year of the most recent regulatory document, from which regulatory threshold levels were derived was compared to the publication year of the studies listed in the ECOTOX database to exclude endpoints, which were more recent than the regulatory documents. This was crucial when the model was set up, because recent studies younger than the regulatory documents cannot be considered in risk evaluations and should therefore not be used in threshold calculations during model calibration and validation, as they would impede a proper analysis of model precision.

Minimum/maximum of concentrations. Filter 12 checks whether a range of endpoint concentrations is specified (i.e., excludes studies which do not report a minimum and maximum concentration). The filter had a substantial power if applied singularly, but was rather considered fine-tuning in additive filter applications, since its application did not result in visible changes of the filter outcome, but improved freshwater model predictions for phenothrin, which previously had underestimated the actual RTL_{reg}. Hence, the filter was not discarded as it still improved model predictions. Further, filter inclusion appears to be relevant since it excludes studies, for which no range of uncertainty for endpoint concentrations was reported and no statistics were calculated. This could indicate that no proper chemical analyses were performed, as a chemical analysis is necessary to provide the measure of uncertainty of the actual concentration, which is essential to provide test statistics.

Definitive endpoints. The inclusion of filter 14 aimed at omitting endpoints which were non-definitive with an operator indicating that the endpoint would be higher or lower than a given concentration. In regulatory documents, this was mostly the case if no LC_{50} or EC_{50} could be determined because the highest concentration tested did not produce an effect of at least 50%. Or, the tests were only run with one concentration and no toxicity was indicated (as typical for tier I testing for aquatic plants and therefore potentially relevant for model extension). A third reason for non-definitive endpoints was that at the limit of solubility no toxic effects (or effects lower than 50%) were reported such that the effect concentration was given as *> level of solubility*, as it was the case for clofentezine. For tebufenozide, in the Preliminary ecological risk assessment (2015, No. 106 in list A2.3, appendix), risk assessors concluded that no risk quotients could be calculated for acute estuarine fish, since no definitive endpoints were reported. Here, the solubility level was close to the highest concentration tested, for which no effects were observed. Therefore, it was assumed that effects from tebufenozide to estuarine fish were unlikely, since estimated environmental concentrations were about 21 times lower than the level of solubility, at which no effects were observable. Also, in the Problem Formulation for indoxacarb, an estuarine acute fish study is provided which reports 40% mortality in the highest concentration tested. Risk assessors argue that in case that no further data was provided, the

conservative assumption of an endpoint equalling the highest tested concentration was used for risk estimation. Thus, EPA's risk assessors handled the reporting of non-definitive endpoints very differently, sometimes omitting values, sometimes accepting values for risk analysis. However, I argue that since non-definitive endpoints give no certainty about a threshold level, only definitive study results were deemed reliable and only studies with definitive endpoints were used for threshold estimations in the present model.

Controls and chemical analyses. Filter 16 and 17 were formulated to filter for studies which included appropriate controls and proper chemical analysis of test concentrations. The singular application of filter 16 excluded 57 and 46%, filter 17 excluded 11 and 10% of endpoints if compared to the null model for freshwater and estuarine models, respectively. Even though both filter criteria did not lead to large changes in remaining endpoints if filters were applied consecutively, the exclusion power of singular filter applications underlines the importance of their inclusion to the RTLE model. This is in line with the rejection rate analysis of the US EPA (1994), which identified issues with controls and chemical analysis as major factors which invalidated study results during the review of studies for ecological risk assessments.

Organisms' characteristics. Filter 18 was formulated to enable the exclusion of distinct organism characteristics, which do not comply with standard test guidelines: For instance, for daphnia tests, 1st instars younger than 24 hours are needed (OCSPP 850.1010, US EPA, 2016*a*, table A6). Therefore, studies with daphnids encoded to be older than 24 hours were excluded. Also, acute fish testing is normally run with certain life stages: According to guideline OCSPP 850.1075 (US EPA, 2016*g*; table A6) juvenile fish shall be used for testing. Here, the filter was formulated to exclude fry, that are early fish life-stages that already resorbed their yolk sac and are freely feeding, even though they are not yet categorized as juveniles (Bone & Moore, 2008; Kendall et al., 1984). This might be the reason, why studies using fry stages of fish might not be used for risk evaluations. In our case, the choice of filter inclusion was mainly data driven, since its application helped to remove too sensitive endpoints, as for fenamiphos. Even though, I cannot tell whether the study was not used by the risk assessors due to the use of the wrong life-stages, or whether another, not-coded reason caused its invalidation, its inclusion did not worsen model precision. Therefore, it was deemed justifiable to keep this filter criteria.

The increase of exclusion power with additive filter applications is paralleled by the loss of data for single substances. In the freshwater model application, for five insecticides no threshold values could be estimated. In the estuarine model application, the number was slightly higher with six insecticides, for which no threshold values could be computed. Hence, the model loses about 10% of substances from the null model to the full model. This loss is probably due to

a low quality of endpoints for the five and six insecticides, such that all endpoints are omitted by the RTLE model. However, it was accepted to rather exclude substances than to include substances with endpoints based on rather unreliable studies. Thus, depending on the quality of the data provided in the ECOTOX database, it is likely that the model does not yield threshold estimates for all substances, even though the amount of substances for which endpoints were yielded was still sufficient as thresholds for about 90% of targeted substances could be obtained.

To summarize, filter criteria selection was mainly driven by the given quality criteria of standard guidelines and the rejection rate analysis of the EPA (as summarized in tables A6-A9). Powerful selection criteria were identified as well as selection criteria that were less powerful in irrelevant endpoint exclusion. The decision, whether criteria were kept in the model or discarded was made case-by-case mainly influenced by the exclusion power of filter steps among all substances, but later, when criteria were formulated for the full model, also to improve the prediction precision for single substances.

Difficulties with filter formulation. Aside the justification of inclusion of filter criteria, there were also reasons to discard irrelevant or not helpful filter criteria. Filter criteria were removed from the model whenever they led to a decrease in model precision or were too strict, as they removed too many relevant datapoints. For instance, there were rejection reasons which were not consistently coded, like the monitoring of physical parameters during the test (temperature, oxygen concentrations), and most entries contained an *NC* (i.e., information was not coded) or *NR* (i.e., information was not reported). This was also the case for important validity criteria as control mortality, which appeared to be only mentioned in the effect comment field. The inclusion of these criteria would have been to strict, and were therefore removed from the model. The problem here is presumably rather the restriction of information provided in the original studies, than the restriction of information which is encoded into the database. Here, the importance of reporting all relevant study features in publications should be emphasized to enable the use of data for risk evaluations.

Further, the encoding of the database did not allow to consider all relevant invalidity criteria during filter formulations (e.g., control mortality), as they are stated in the standard test guidelines. Here, especially the *effect comment* parameter contained information which could have led to a rating of study results to be invalid, but since they were coded non-standardized in free texts, the directed formulation of filter criteria was hindered, as discussed above.

In addition, according to test guidelines, at least five levels of concentrations should be tested in definitive tests. However, during filter formulation, it was not possible to further refine the model since in the full model, for all remaining studies no information on the number of tested

concentrations was provided in the database (i.e., *NR*, not reported). According to the ECOTOX coding guidelines (p. 32, US EPA, 2016*j*), if a range of concentrations is not given, it can be assumed that these studies contained calculated endpoints for which no range of concentrations was specified. It appears to be the same reason why it was not possible to filter for the number of test organisms per treatment, as numbers were not reported for the remaining endpoints after full model application. Thus, it is not possible to filter for the specific numbers of test organisms, which shall be used per test concentrations, as well as numbers of replicates, even though both might have been valuable filter criteria since the test standard guidelines report explicit numbers of replicates and test organisms per concentration to be used in standard tests. However, the absence of information for studies that pass the full model criteria might imply that to some extent, studies are used for risk assessments for which no raw data is available. This would mean that risk assessments were based on studies that report only calculated endpoints, such that results and calculations might not be completely traceable. However, this goes beyond the scope of the present study and should be addressed in further research.

Finally, the formulation of exclusion criteria might also be hampered due to the fact, that the application of filter criteria does not allow for expert judgements. As some study validity criteria are very strict and lead to the direct exclusion of the study if they are violated, others are more flexible, as studies might be accepted for risk evaluations, even though one or more validity criteria are not met. Here, experts could argue that results are still robust enough such that studies would be rated as reliable, for example to avoid unnecessary animal testing if additional studies are not thought to result in a gain of knowledge on the toxicity of a substance. Unfortunately, for the RTLE model, there is no possibility to incorporate the expert judgement, as filter criteria are applied definitely. However, even though the model does not allow case-by-case decisions in terms of that sometimes validity criteria lead to the exclusion of a study and sometimes, studies might still be found sufficient although a criterion is not passed, the RTLE model performs very good. How good the model performs and whether there are restrictions for its application, is discussed in the next section.

4.2 Model performance

As discussed in the section above, the formulated filter criteria resulted in three models, which differed in their strictness. The evaluation of model performances indicates that the full model yields the most precise predictions applying estuarine and freshwater models. The application of selection criteria was that powerful (i.e., excluded a sufficient number of invalid, low endpoints) such that it was reasonable to take the minimum value of each endpoint distribution per substance as basis for threshold estimations in both freshwater and estuarine model applications.

Thus, the RTLE model allows to follow the rules stated in ecological risk assessment guidelines (US EPA, 2015*a*), which consider the most sensitive reliable endpoint for risk quotient calculations. In freshwater full model application, about one third of predictions were correct (30.6% (bootstrapped 95%-CI 20.8 - 41.7%)), and the likelihood that RTL_es undermatch the RTL_{reg} was low. However, the likelihood to overshoot the RTL_{reg} was still fairly high, as the median of all of RTL_es exceeds the respective RTL_{reg}s by an order of magnitude of 1.7, while 95% of RTL_es did not exceed an order of magnitude of 3.5 (fig. 3). The estuarine model performed even better, with regard to the higher proportion of 43.5% (bootstrapped 95%-CI 32.3 - 54.8%) for which predictions were correct. However, the proportion of RTL_es which underestimated the RTL_{reg} was slightly higher in estuarine model predictions (16.1% (bootstrapped 95%-CI 8.1 - 25.8%) versus 12.5% (bootstrapped 95%-CI 5.6 - 20.8%), for estuarine and freshwater models, respectively) what relativizes the overall better performance of the estuarine model slightly, even though the difference between the proportions that underestimated the RTL_{reg}s was not statistically significant (fig. 13, table A1, appendix). The extent to which the RTL_{reg}s were overestimated by the RTL_e was also lower in estuarine model applications, as the median exceeded the respective RTL_{reg}s by an order of magnitude of 0.8 (compared to 1.7 for freshwater predictions), while 95% of RTL_es did not exceed an order of magnitude of 3.7 (compared to 3.9 for freshwater predictions, fig. 3). The obtained orders of magnitudes of exceedances in freshwater and estuarine model applications were high, even though if compared to the exceedance of environmental concentrations of threshold values, exceedances are similar to the values reported by Schulz (2004), such that even though RTL_es tend to largely overestimate the RTL_{reg}, still a risk could be identified if compared to environmental concentrations. However, this applies only to large environmental concentrations, as for lower concentrations, which do not exceed the RTL_e but the RTL_{reg}, the actual risk would not be identified if compared to estimated thresholds. Therefore, at this point no generalized conclusions can be drawn on the sensitivity of RTL_es, because other studies and measurements report different degrees of exceedances. For instance, the study by Larson et al. (1999, from visual inspection) measured concentrations that exceed threshold levels by 1-2 orders of magnitude. Here, RTL_es exceeding the respective RTL_{reg}s by the orders of magnitude observed in the present study would not be tolerable to indicate the existing risk, which would be identified if actual threshold values were used. However, an extensive evaluation of the sensitivity of the estimated RTL_es in comparison to their respective RTL_{reg}s (i.e., the likelihood to indicate a risk applying RTL_es, if RTL_{reg}s were exceeded) if compared to environmental concentrations could not yet be conducted and should therefore be subject to future research.

Model precision could further be enhanced by only considering substances for which the

endpoint relevant for risk assessment was included to the database: full model predictions for endpoints included to the ECOTOX database performed significantly better if compared to the endpoints which were not included (fig. 14). Here, a large proportion of threshold levels cannot be predicted correctly, since the actual regulatory endpoints are missing in the database. This might be due to the fact that only part of the studies used in risk assessments are published in the open literature. Instead, registrant-submitted studies remain mostly un-published and non-peer-reviewed, as they are conducted in contract laboratories, which provide study reports including the data submitted for registration. Here, the inclusion of a second database, the OPP Pesticide Toxicity database (http://www.ipmcenters.org/ecotox/), might aid to improve predictions, as it contains EPA's reviewed ecotoxicity data of registered and formerly registered pesticides. It might be interesting to check whether a combination of both databases yields better predictions, since the OPP database is thought to include the missing endpoints from non-published studies, that are not available in the ECOTOX database. However, this analysis is beyond the scope of the present project, but should be evaluated in a future project. It remains that the proportion of substances for which the RTL_{reg} was underestimated was not statistically different from the actual full model predictions with regard to the presence or absence of data in the ECOTOX database (fig. 14). Consequently, the RTLE model is also applicable for substances for which the correct endpoint is not included in the database. For these substances, but also for the substances for which an actual RTL_{reg} is included in the database, it needs to be considered for further interpretations that the estimated RTL_es tend to overestimate the respective RTL_{reg}, what results in rather liberal threshold estimates if used as a basis for risk evaluations of for instance environmental concentrations. Thus, it could be the case, that threshold estimates indicate no risk if compared to environmental concentrations, where there might actually occur a risk to non-target organisms. As proposed in the section above, a comparison between RTL_es and RTL_{reg}s with regards to the predicted risks if compared to environmental concentrations, could help to evaluate the influence of the degree of overestimations on model performance. In turn, it is not likely that the RTLE model indicates a risk, if there is none, if threshold levels are compared to environmental concentrations. This means, if threshold levels generated with the RTLE model are exceeded, it is very likely that the protection goals of risk assessments cannot be met and detrimental effects on non-target organisms cannot be excluded. In conclusion, model precision is fairly good, as at least the likelihood to falsely indicate a risk if there is non, is low independently of whether the respective RTL_{reg}s for a substance is included in the database or not.

As a second step of model validation, it was analyzed whether there were correlations or linear relationships between external factors and the generated threshold estimates. The RTL_{reg}s

were strongly correlated with the RTL_es (table 2-3). Correlation coefficients were high with 0.83 (95%-CI 0.74 - 0.89) and 0.9 (95%-CI 0.84 - 0.93) indicating a strong linear relationship between the two parameters in freshwater and estuarine model applications, respectively. Correlation coefficients are in line with the evaluation of model performance: A perfect correlation (r = 1) would be achieved if all datapoints lay on one line without points scattering around this line. Since scatter was lower (r closer to 1) in the estuarine model predictions, it is indicated that the precision of the RTLE model was slightly better for estuarine applications, as already seen in the evaluation of model precision above.

The comparison between RTL_{reg} and RTL_e with the ratio $\log_{10}(RTL_e/RTL_{reg})$ was conducted to test whether higher or lower endpoints were easier or more difficult to predict, and whether higher or lower estimates resulted from a higher or lower model precision (table 2-3; fig. A1-2, appendix). For freshwater model application, the model precision was not dependent on the RTL_e, but for the regulatory thresholds, a significant decrease of model precision with an increase of the RTL_{reg} was apparent. The linear relationship is mainly influenced by a group (n = 5) of very low threshold values, which were more likely to be overestimated. This group comprised five pyrethroids, that were already identified as being clustered when RTL_es were plotted against RTL_{reg}s, as briefly discussed below. In contrast, there were only two influential points with a low model precision, but high RTL_{reg}s (i.e., comparably lower toxicity) thereby influencing the indicated linear relationship substantially. For estuarine model application, no such correlation was apparent. And, the correlations between RTL_es and the ratio $\log_{10}(RTL_e/RTL_{reg})$ for freshwater and estuarine model applications were also low, even though for estuarine model application, the correlation was influenced by three leverage points, leading to a statistically, but not meaningful, significant correlation. In summary, the model precision is deemed sufficiently, as the model estimates were not better or worse in dependence of the actual RTL_{reg} or RTL_e, except for the the group of pyrethroids, for which, however, RTL_es are overpredicted and not underpredicted, such that this imprecision is suspected to be tolerable.

The test for correlations with physical-chemical parameters, did not reveal severe dependencies to account for. Aside the test for relationships between the estimated and actual threshold levels, the most important question to be addressed was whether the precision of predictions was dependent on or interrelated with external parameters: In all tested cases except for one, the dependencies detected for the RTL_{reg} or RTL_e did not translate into the measure of model precision. The only association was found in the freshwater model, where the ratio $\log_{10}(RTL_e/RTL_{reg})$ was significantly correlated with the molecular weight of substances (0.3, 95%-CI 0.03 - 0.52, p = 0.003, df = 52, t = 2.23). However, the confidence interval was large including 0.03, what would be close to 0 indicating no association, and 0.52, what would indi-

cate a strong relationship. Judged by the size of this confidence interval, considerable uncertainty remains with regards to correlation coefficient even though it was statistically significant. Therefore, the indicated correlation between the ratio $\log_{10}(RTL_e/RTL_{reg})$ and molecular weight should not be over-interpreted or overweighted. Most important, since only a low number of predictions underestimated the RTL_{reg}, and the indicated correlation was positively with most scattered datapoints exceeding 0, the likelihood to underestimate the RTL_{reg} did not depend on the molecular weight of the substance for which threshold values were estimated. Therefore, it appears as if there is no need to correct the model for any imprecision with regards to the molecular weight.

Further, the partially observed associations between the RTL_{reg}s / RTL_es and the solubility and the partitioning coefficient did not translate into the comparison of physical-chemical parameters to model precision. Even though there are some low diffuse patterns for both models evident, which are rather contradictious if models are compared, no clear patterns are apparent. In addition, even though the degree of association between the reference year and the ratio $\log_{10}(RTL_e/RTL_{reg})$ in the freshwater model application, and the degree of association between the number of endpoints and the ratio $\log_{10}(RTL_e/RTL_{reg})$ for the estuarine model application were significant, the indicated association did not reveal a relationship between the parameters and the likelihood to underestimate the RTL_{reg}, which was apparent from visual inspections. Instead, the lines of best fits were rather influenced by the scatter in the datapoints which overestimated the RTL_{reg}. However, since no clear patterns evolved and correlations could mainly be attributed to leverage points, it is not deemed necessary to correct the model for any of the tested external parameters.

The visual inspection of potential patterns in prediction precision for the different substance groups did not reveal any difficulties in model performance as there were no groups for which RTL_es were underestimated. However, if RTL_es were plotted as a function of RTL_{reg}s it appears as if some pyrethroids were clustered (fig. A8-10) as they tended to be overestimated. This was probably due to the newly published, preliminary regulatory document (No. 29 of list A2.3), for which the new regulatory endpoints were not yet included to the ECOTOX database. Finally, since in the calibration and validation datasets RTL_{reg}s were equally over- or underestimated, there was no need to adapt the model to account for potential differences.

4.3 Model application

The application of the RTLE model for a set of insecticides, for which no regulatory documents could be retrieved, resulted in RTL_es for six substances. All of the insecticides tested in model application where no longer permitted, such that no new literature searches are performed,

what might reduce the number of available studies in the ECOTOX database. However, for scientific risk evaluations of the current state of pollution of aquatic environments, it might be still interesting to compare concentrations of chemicals, which are no longer applied, but are still traceable in ecosystems. Here, for instance, Hoffman et al. (2000) did not find a water quality threshold value for fonofos, such that the measured concentrations in their study could not be related to a benchmark. In such cases, the RTLE model could aid, as the derived RTL_es could be used to fill data-gaps of missing threshold values.

Unfortunately, the yield of RTL_es was very low, as only six out of 29 insecticides yielded threshold estimates. This could also be due to the fact that older studies do not comply to the applied validity criteria and are thus removed from the results as filters are applied. Again, the OPP database could aid, as it is supposed to provide also endpoints of formerly registered substances. Hence, the inclusion of the OPP database into the RTLE model is expected to improve the application spectrum of the model, such that threshold equivalents could be computed for a higher number of substances, for which no regulatory documents are available.

The test of the transferability of the RTLE model to both herbicides and fungicides yielded promising results, as the original model could be easily applied to fungicides without obtaining significantly worse threshold estimates. And, slight adaptions of the filter formulations enabled predictions for herbicides of equal precision as the actual model for insecticides, for which the model had originally been calibrated. Filter criteria to extent the model for herbicides and fungicides were mainly of technical nature, as filters were re-formulated to allow aquatic plant species and to select the respective endpoints (e.g., growth) and study durations (e.g., 7 days for *Lemna sp.* tests). However, since the exposure of herbicides and fungicides in the environment follows different patterns, further adaptions would be necessary to enable the prediction of chronic threshold level equivalents, as discussed in the following section.

4.4 Study perspectives and model assumptions

The perspective of the study was to test whether the ECOTOX database can be used to estimate threshold level equivalents. As successfully introduced and demonstrated above, the for this purpose developed RTLE model is a promising tool to estimate threshold level equivalents for pesticides. The first assumption that, the most sensitive endpoint per substance is relevant for RTLe computations, holds true, since the applied filter criteria were that good that the most sensitive endpoint per substance remaining after filter applications described best the approximated position of the actual RTL_{reg}. The second assumption targeted at justifying the use of acute endpoints for RTL_e estimation to be used in risk evaluations, as done by Stehle & Schulz (2015*a*). Since insecticides occur in irregular short peaks in freshwaters (Gilliom et al.,

2006; Stehle et al., 2013), the second assumption is valid and it is plausible to base to RTLE model on acute toxicity data, as the derived RTL_es are to be used to evaluate potential risks to non-target organisms exposed to insecticides in the field. In contrast, to enable the model application for fungicides and herbicides, the solely consideration of acute toxicity data might not be adequate. Even though the extension of the model for herbicides and fungicides worked well for the acute data, to account for realistic exposure scenarios, an extension of the RTLE model would need to be calibrated with chronic endpoints, since exposure in the field is better described by longer exposure periods (Gilliom et al., 2006; Kreuger, 1998; Kreuger & Brink, 1988; Schreiner et al., 2016). This is especially important, since chronic endpoints are potentially lower than acute endpoints (acute to chronic ratio > 1; Länge et al., 1998), such that the estimation of chronic threshold levels could result in more sensitive RTL_es, even though a higher level of concern would be applied (LOC = 1). The comparison of lower threshold levels to actual exposures would be more sensitive to indicating a risk than if acute threshold estimates were used. Further, if thresholds are based on chronic toxicity data, the risk evaluation becomes more realistic, as not only lethal effects are considered, but environmental concentrations can be related to concentrations, at which sublethal effects as the impairment of reproduction or growth could occur. Thus, to account for more realistic types of exposures and effects, an extension of the model to account for chronic endpoints is crucial. The third assumption, that aquatic plant endpoints are not relevant for RTLE model applications is in line with the most sensitive endpoints reported in regulatory documents for the calibration and validation datasets for insecticides (table A3 and A4, appendix). For fungicides, plant endpoints are not necessarily indispensable, since the original model predictions were not significantly different from the adapted model predictions if applied for fungicides, even though the adapted model performed slightly better for freshwater model application, if aquatic plants were considered. In contrast, for herbicides, the inclusion of aquatic plants improved model precision significantly. However, the rate of underpredictions of the actual RTL_{reg}s was independently of whether aquatic plants were considered or not in fungicide and herbicide model extensions, and equally low as for the calibrated and validated insecticide model. Therefore, the third assumption is not violated for insecticide RTLE model applications, but aquatic endpoints shall rather be considered if the model is extended to decrease the degree to which RTL_{reg}s are overestimated.

To sum up, aside a few restrictions or identified needs for further adaptions in the present analysis of model performance (e.g., high likelihood to overestimate actual RTL_{reg}s, missing entries of endpoints in the ECOTOX database, no expert judgement, loss of substances for which no RTL_es can be provided), the introduced RTLE model proved to be a good tool for threshold level estimations for insecticides, whereby the application for fungicides and herbicides also

works fine, but would need some additional extensions to consider chronic endpoints. Further, the model successfully estimated threshold values for substances for which no regulatory documents were available. Here, the model extension to include data provided in the OPP database is expected to further enhance the power of the RTLE model. In future, the RTLE model could provide access to a high number of regulatory threshold level equivalents to a broader public of researchers, as its application might be especially interesting for ecotoxicologists, who are not familiar with regulations or willing to review regulatory documents to retrieve endpoints, but want to include threshold values into their analyses. The application of the RTLE model saves lots of time and resources (i.e., manpower), that would be needed if RTL_{reg}s were to be collected manually. Thus, the availability of the new RTLE model shows great promise for its future application in risk evaluations. A deeper discussion on how the RTLE model could be used in risk evaluations is provided below.

4.5 Future perspective - Application to scientific evaluations of monitoring data

As a concluding part of the present study, a possible future application of the newly developed RTLE model should be highlighted. The RTLE model was developed to enable a comprehensive risk evaluation by comparing the estimated RTL_es to environmental concentrations. The US EPA provides a large dataset of monitoring data (WQP, 2018). To analyze the impact of environmental concentrations on the biota, exposure needs to be linked to effects. However, a comprehensive analysis of this monitoring data was yet hindered by the available number of regulatory threshold values, such that up to now, most risk evaluations were restricted to a limited amount of substances included to the evaluations.

In the past, huge efforts were made to draw a comprehensive picture of pesticide pollutions and its effects. In the US, governmental pesticide monitoring data is freely accessible, and different studies are available relating these monitored environmental concentrations to water quality criteria (Hoffman et al., 2000), regulatory threshold levels and other benchmarks (Gilliom et al., 2006), chronic aquatic life benchmarks (Stone et al., 2014), and regulatory threshold levels as well as non-regulatory guideline concentrations (Larson et al., 1999). However, the analysis of pesticides in US surface waters over the last 20 years does not yield a comprehensive picture of the degree of insecticide pollution, either because the number of assessed insecticides was low (n = 13; Hoffman et al.,2000) or the provided water quality benchmarks or the list of threshold values for analyzed insecticides is incomplete (Larson et al., 1999).

For Europe, Malaj et al. (2014) analyzed the influence of 223 organic chemicals, of which 52 were insecticides at about 4000 European monitoring sites. Here, threshold values were derived by applying safety factors (e.g., 10, 100, 1000) to standard test species (algae, daph-

nia, fish), to obtain acute and chronic benchmarks. However, toxicity data was obtained from multiple sources, whereby, it was not explained, how the final toxicity value was selected, if, for instance, more than one endpoint per species were available. Here, Stehle & Schulz provided in 2015*b* the first european analysis of risks of pesticide concentrations in surface waters in agricultural dominated environments that was based on actual regulatory threshold values. Here, 1554 reported measurements compiled from peer-reviewed literature were compared to RACs (regulatory acceptable concentrations) of 23 different insecticides. In 2017, a detailed analysis of pesticides in German surface waters was evaluated by Szöcs et al., who compared routine monitoring data from grab sampling to a list of 107 RACs provided by the German Environment Agency (UBA). Their evaluation included also herbicides and fungicides, such that only 22 of the substances were insecticides, with a partly different spectrum of insecticides and respective RACs analyzed if compared to the substances and RACs used by Stehle & Schulz (2015*b*).

The only global assessment was published by Stehle & Schulz (2015*a*), in which the exposure of non-target organisms was related to regulatory threshold levels of insecticides considering freshwater and estuarine systems. Here, in an extensive literature review, measured insecticide concentrations were compiled from peer-reviewed studies, and compared to 28 regulatory threshold values of insecticides. Regulatory threshold values were used for the US, EU and Canada, whereby for the remaining regions of this global assessment, mean values of EU and US thresholds were calculated. Irrespective of the different scales, type of threshold levels and substance spectra, all studies had in common that threshold values were regularly exceeded (Gilliom et al., 2006; Hoffman et al., 2000; Larson et al., 1999; Malaj et al., 2014; Stehle & Schulz, 2015*a*,*b*; Stone et al., 2014; Szöcs et al., 2017), highlighting the relevance of comprehensive evaluations of pesticide concentrations in aquatic systems. This list of reviewed studies indicating risks of pesticides to freshwater ecosystems is however not comprehensive and could be extended for studies focusing more on the types of effects induced by concentrations even below regulatory threshold values on different scales (e.g., see Schäfer et al., 2012; Beketov et al., 2013). The focus in the present study lay more on the different threshold values that were related to environmental concentrations.

In the presented studies, it appears as if the amount of analyzed substances was rather hindered by the amount of available threshold values, than by the amount of measured substances. This was especially apparent for the studies using the WQP dataset to evaluate pesticide contaminations in the US. Since the US EPA does not provide a list of regulatory threshold levels as the German UBA does (even though the UBA does not provide freely accessible monitoring data), and the computation of actual RTLs is cumbersome and time consuming due to the unstructured provision of documents, as discussed above, the analysis of the WQP dataset

was yet restricted to the amount of available threshold values. Here, the present study could prove useful in several ways: besides providing the retrieved list of actual 83 and 73 RTL_{reg}s for insecticides, 27 and 8 RTL_{reg}s for herbicides, as well as 15 and 12 RTL_{reg}s for fungicides, for estuarine and freshwater systems, respectively, the model also has the potential to generate an even higher number of threshold values through the RTLE model application. As introduced above, the RTLE model could thereby serve as a handy tool to estimate threshold levels for comprehensive risk evaluations of insecticide concentrations in the environment. And, if further adapted, the RTLE model is likely to yield RTL_es for an even broader substance spectrum including acute and chronic RTL_es for fungicides and herbicides. Then, the application of the model allows for an extensive analysis of monitoring data for a broad number of substances to gain a comprehensive picture when the risk of pesticides in the environment is evaluated for the US, thereby contributing to draw the bigger picture of global pesticide pollutions of surface waters.

Bibliography

Auguie, B. (2017), *gridExtra: Miscellaneous Functions for "Grid" Graphics*, Available at https://CRAN.R-project.org/package=gridExtra.

Beketov, M. A., Kefford, B. J., Schäfer, R. B. & Liess, M. (2013), 'Pesticides reduce regional biodiversity of stream invertebrates', *Proceedings of the National Academy of Sciences* **110**(27), 11039–11043.

Bereswill, R., Streloke, M. & Schulz, R. (2013), 'Current-use pesticides in stream water and suspended particles following runoff: Exposure, effects, and mitigation requirements', *Environmental Toxicology and Chemistry* **32**(6), 1254–1263.

Bone, Q. & Moore, R. (2008), Reproduction and life histories, *in* 'Biology of Fishes', 3 edn, Taylor & Francis, pp. 217–254.

Brady, D. J. (2011), 'Evaluation guidelines for ecological toxicity data in the open literature', *Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/evaluation-guidelines-ecological-toxicity-data-open Accessed March 3, 2018* .

Campbell, M. J., Daly, L. E. & Machin, D. (2005), Special topics, *in* D. G. Altman, D. Machin, T. N. Bryant & M. J. Gardner, eds, 'Statistics with confidence: confidence intervals and statistical guidelines', 2. ed. edn, BMJ Books, London, pp. 153–167. OCLC: 838006135.

Campbell, M. J. & Gardner, M. J. (2005), Medians and their differences, *in* D. G. Altman, D. Machin, T. N. Bryant & M. J. Gardner, eds, 'Statistics with confidence: confidence intervals and statistical guidelines', 2. ed. edn, BMJ Books, London, pp. 36–44. OCLC: 838006135.

Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S. K. & Tiffin, N. (2016), 'RPostgreSQL: R interface to the PostgreSQL database system', *Available at https://CRAN.R-project.org/package=RPostgreSQL* .

Covich, A. P., Ewel, K. C., Hall, R. O., Giller, P. E., Goedkoop, W. & Merritt, D. M. (2004), Ecosystem services provided by freshwater benthos, *in* D. H. Wall, ed., 'Sustaining Biodiversity and Ecosystem Services in Soils and Sediments', Vol. 64, Island Press, pp. 45–72.

Cumming, G. & Finch, S. (2005), 'Inference by eye: confidence intervals and how to read pictures of data.', *American Psychologist* **60**(2), 170–180.

Dowle, M. & Srinivasan, A. (2017), *data.table: Extension of 'data.frame'*, Available at https://CRAN.R-project.org/package=data.table.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J. & Sullivan, C. A. (2006), 'Freshwater biodiversity: importance, threats, status and conservation challenges', *Biological Reviews* **81**(2), 163–182.

FAO (2017), 'FAOSTAT Database. Food and Agriculture Organization of the United Nations (FAO).', *Available at http://www.fao.org/faostat/en/#data/EL Accessed April 4, 2018* .

FIFRA (1947), 'Federal insecticide, fungicide and rodenticide act', *U.S. Federal Law, amended in 1972 and 1988* .

Finizio, A. & Villa, S. (2002), 'Environmental risk assessment for pesticides: A tool for decision making', *Environmental Impact Assessment Review* **22**(3), 235–248.

Fox, J. & Weisberg, S. (2011), 'An R Companion to Applied Regression', *Available at http://socserv.socsci.mcmaster.ca/jfox/Books/Companion* .

Gama, S., Arnot, J. A. & Mackay, D. (2012), Toxic Organic Chemicals, *in* J. S. Gulliver, ed., 'Transport and Fate of Chemicals in the Environment', Springer, pp. 41–63.

Gianessi, L. & Reigner, N. (2006), 'Pesticide Use in U.S. Crop Production: 2002', *CropLife Foundation, Washington D.C.* pp. 1–36.

Giesy, J. P., Solomon, K. R., Coats, J. R., Dixon, K. R., Giddings, J. M. & Kenaga, E. E. (1999), Chlorpyrifos: Ecological Risk Assessment in North American Aquatic Environments,

in G. W. Ware, ed., 'Reviews of Environmental Contamination and Toxicology: Continuation of Residue Reviews', Vol. 160, Springer New York, New York, NY, pp. 1–129.

Gilliom, R. J., Barbash, J. E., Crawford, C. G., Hamilton, P. A., Martin, J. D., Nakagaki, N., Nowell, L. H., Scott, J. C., Stackelberg, P. E., Thelin, G. P. & others (2006), 'The quality of our nation's waters - pesticides in the nation's streams and ground water, 1992–2001: U.S. geological survey circular 1291'.

Hoffman, R. S., Capel, P. D. & Larson, S. J. (2000), 'Comparison of pesticides in eight US urban streams', *Environmental Toxicology and Chemistry* **19**(9), 2249–2258.

Kegley, S. E., Hill, B. R., Orme, S. & Choi, A. H. (2016), 'PAN Pesticide Database, Pesticide Action Network, North America (Oakland, CA)', *Available at http:www.pesticideinfo.org* .

Kendall, A., Ahlstrom, E. & Moser, H. (1984), Early life history stages of fishes and their characters, *in* H. Moser, W. Richards, D. Cohen, M. Fahay, A. Kendall & S. Richardson, eds, 'Ontogeny and systematics of fishes', American Society of Ichthyologists and Herpetologists, pp. 11–22.

Kreuger, J. (1998), 'Pesticides in stream water within an agricultural catchment in southern Sweden, 1990–1996', *Science of the Total Environment* **216**(3), 227–251.

Kreuger, J. K. & Brink, N. (1988), Losses of pesticides from agriculture, *in* 'Pesticides: Food and environmental implications.', International Atomic Energy Agency, Vienna, pp. 101–112.

Länge, R., Hutchinson, T. H., Scholz, N. & Solbé, J. (1998), 'Analysis of the ECETOC aquatic toxicity (EAT) database II—Comparison of acute to chronic ratios for various aquatic organisms and chemical substances', *Chemosphere* **36**(1), 115–127.

Larson, S. J., Gilliom, R. J. & Capel, P. D. (1999), Pesticides in streams of the United States: Initial results from the National Water-Quality Assessment Program. US Geological Survey USGS WRIR 98-4222. U.S. Geological Survey, Sacramento, CA.

Liess, M. & Schulz, R. (1999), 'Linking insecticide contamination and population response in an agricultural stream', *Environmental Toxicology and Chemistry* **18**(9), 1948–1955.

Malaj, E., Peter, C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., Brack, W. & Schäfer, R. B. (2014), 'Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale', *Proceedings of the National Academy of Sciences* **111**(26), 9549–9554.

Matson, P. A., Parton, W. J., Power, A. G. & Swift, M. J. (1997), 'Agricultural intensification and ecosystem properties', *Science* **277**(5325), 504–509.

Millennium Ecosystem Assessment (2005), 'Ecosystems and human well-being: synthesis', *Island, Washington, DC* .

Müller, K., Bach, M., Hartmann, H., Spiteller, M. & Frede, H.-G. (2002), 'Point-and nonpoint-source pesticide contamination in the Zwester Ohm catchment, Germany', *Journal of Environmental Quality* **31**(1), 309–318.

OECD (2004*a*), *Test No. 202: Daphnia sp. Acute Immobilisation Test*, Available at https://www.oecd-ilibrary.org/content/publication/9789264069947-en Accessed April 4, 2018.

OECD (2004*b*), *Test No. 218: Sediment-Water Chironomid Toxicity Using Spiked Sediment*, Available at https://www.oecd-ilibrary.org/content/publication/9789264070264-en Accessed April 4, 2018.

OECD (2011), *Test No. 235: Chironomus sp., Acute Immobilisation Test*, Available at https://www.oecd-ilibrary.org/content/publication/9789264122383-en Accessed April 4, 2018.

Oerke, E.-C. (2006), 'Crop losses to pests', *The Journal of Agricultural Science* **144**(1), 31–43.

Pelaez, V., da Silva, L. R. & Araújo, E. B. (2013), 'Regulation of pesticides: A comparative analysis', *Science and Public Policy* **40**(5), 644–656.

Pruim, R., Kaplan, D. T. & Horton, N. J. (2017), 'The mosaic Package: Helping Students to 'Think with Data' Using R', *The R Journal. Available at https://journal.r-project.org/archive/2017/RJ-2017-024/index.html* **9**(1), 77–102.

R Core Team (2016), 'R: A Language and Environment for Statistical Computing', *Available at https://www.R-project.org/* .

Sanchez-Bayo, F. P. (2012), 'Insecticides mode of action in relation to their toxicity to non-target organisms', *Journal of Environmental & Analytical Toxicology* **S4**(002), 1–11.

Schäfer, R. B., Caquet, T., Siimes, K., Mueller, R., Lagadic, L. & Liess, M. (2007), 'Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe', *Science of the Total Environment* **382**(2-3), 272–285.

Schäfer, R. B., von der Ohe, P. C., Rasmussen, J., Kefford, B. J., Beketov, M. A., Schulz, R. & Liess, M. (2012), 'Thresholds for the effects of pesticides on invertebrate communities and leaf breakdown in stream ecosystems', *Environmental Science & Technology* **46**(9), 5134–5142.

Schreiner, V. C., Szöcs, E., Bhowmik, A. K., Vijver, M. G. & Schäfer, R. B. (2016), 'Pesticide mixtures in streams of several European countries and the USA', *Science of The Total Environment* **573**, 680–689.

Schulz, R. (2004), 'Field studies on exposure, effects, and risk mitigation of aquatic nonpoint-source insecticide pollution', *Journal of Environmental Quality* **33**(2), 419–448.

Stehle, S., Knäbel, A. & Schulz, R. (2013), 'Probabilistic risk assessment of insecticide concentrations in agricultural surface waters: a critical appraisal', *Environmental Monitoring and Assessment* **185**(8), 6295–6310.

Stehle, S. & Schulz, R. (2015*a*), 'Agricultural insecticides threaten surface waters at the global scale', *Proceedings of the National Academy of Sciences* **112**(18), 5750–5755.

Stehle, S. & Schulz, R. (2015*b*), 'Pesticide authorization in the EU—environment unprotected?', *Environmental Science and Pollution Research* **22**(24), 19632–19647.

Stone, W. W., Gilliom, R. J. & Ryberg, K. R. (2014), 'Pesticides in U.S. Streams and Rivers: Occurrence and Trends during 1992–2011', *Environmental Science & Technology* **48**(19), 11025–11030.

Szöcs, E. (2016), 'Build a local version of the EPA ECOTOX database', Available at http://edild.github.io/localecotox/ Accessed November 8, 2017.

Szöcs, E., Brinke, M., Karaoglan, B. & Schäfer, R. B. (2017), 'Large scale risks from agricultural pesticides in small streams', *Environmental Science & Technology* **51**(13), 7378–7385.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W. H., Simberloff, D. & Swackhamer, D. (2001), 'Forecasting agriculturally driven global environmental change', *Science* **292**(5515), 281–284.

Touart, L. W. & Maciorowski, A. F. (1997), 'Information needs for pesticide registration in the United States', *Ecological Applications* **7**(4), 1086–1093.

US EPA (1994), 'Pesticide Reregistration Rejection Rate Analysis: Ecological Effects', *NEPIS. Available at https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=20000F9S.txt Accessed April 4, 2018* pp. 1–208.

US EPA (1998), 'Guidelines for Ecological Risk Assessment', *Federal Register, Washington, DC.* **63**(93), 26846–26924.

US EPA (2007), 'Special Docket for Pesticide Reregistration Risk Assessments - Risk assessments and related documents (DVD Set)', *Available at https://www.regulations.gov/docket?D=EPA-HQ-OPP-2007-0151* .

US EPA (2012*a*), 'Ecological Effects Test Guidelines OCSPP 850.4400: Aquatic Plant Toxicity Test Using Lemna spp.', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0027 Accessed April 4, 2018* pp. 1–24.

US EPA (2012*b*), 'Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0003 Accessed April 4, 2018* pp. 1–28.

US EPA (2012*c*), 'Ecological Effects Test Guidelines OCSPP 850.4550: Cyanobacteria (Anabaena flos-aquae) Toxicity', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0004 Accessed April 4, 2018* pp. 1–26.

US EPA (2013), 'Pesticide Registration Manual: Introduction', *Available at https://www.epa.gov/pesticide-registration/pesticide-registration-manual-introduction Accessed April 4, 2018* .

US EPA (2015*a*), 'Technical Overview of Ecological Risk Assessment - Analysis Phase: Ecological Effects Characterization', *Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/technical-overview-ecological-risk-assessment-0 Accessed April 4, 2018* .

US EPA (2015*b*), 'Technical Overview of Ecological Risk Assessment - Analysis Phase: Exposure Characterization', *Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/technical-overview-ecological-risk-assessment-1 Accessed April 4, 2018* .

US EPA (2015*c*), 'Technical Overview of Ecological Risk Assessment: Risk Characterization', *Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/technical-overview-ecological-risk-assessment-risk Accessed April 4, 2018* .

US EPA (2016*a*), 'Ecological Effects Test Guidelines OCSPP 850.1010: Aquatic Invertebrate Acute Toxicity Test, Freshwater Daphnids', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0041 Accessed April 4, 2018* pp. 1–17.

US EPA (2016*b*), 'Ecological Effects Test Guidelines OCSPP 850.1020: Gammarid Amphipod Acute Toxicity Test', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0040 Accessed April 4, 2018* pp. 1–18.

US EPA (2016*c*), 'Ecological Effects Test Guidelines OCSPP 850.1025: Oyster Acute Toxicity Test (Shell Deposition)', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0039 Accessed April 4, 2018* pp. 1–19.

US EPA (2016*d*), 'Ecological Effects Test Guidelines OCSPP 850.1035: Mysid Acute Toxicity Test', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0038 Accessed April 4, 2018* pp. 1–17.

US EPA (2016*e*), 'Ecological Effects Test Guidelines OCSPP 850.1045: Penaeid Acute Toxicity Test', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0037 Accessed April 4, 2018* pp. 1–18.

US EPA (2016*f*), 'Ecological Effects Test Guidelines OCSPP 850.1055: Bivalve Acute Toxicity Test (Embryo-Larval)', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0036Accessed April 4, 2018* pp. 1–18.

US EPA (2016*g*), 'Ecological Effects Test Guidelines OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test', *Available athttps://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0035 Accessed April 4, 2018* pp. 1–19.

US EPA (2016*h*), 'Ecological Effects Test Guidelines OCSPP 850.1735: Spiked Whole Sediment 10-Day Toxicity Test, Freshwater Invertebrates', *Available at https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0154-0030 Accessed April 4, 2018* pp. 1–35.

US EPA (2016*i*), 'Procedures for Encoding Toxicity Data Published in the Open Literature for Use in Ecological Risk Assessment - EFED Coding Guidelines', *Available at https://cfpub.epa.gov/ecotox/blackbox/help/EFEDCodingGuidelines.pdf Accessed April 4, 2018* pp. 1–38.

US EPA (2016*j*), 'Unify Toxicity Module ECOTOXicology Knowledgebase System ECOTOX Coding Guidelines', *Available at https://cfpub.epa.gov/ecotox-new/pdf/ecotoxsop.pdf Accessed April 4, 2018* pp. 1–84.

US EPA (2017*a*), 'Aquatic Life Benchmarks and Ecological Risk Assessments for Registered Pesticides', *Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/aquatic-life-benchmarks-and-ecological-risk Accessed April 4, 2018* .

US EPA (2017*b*), 'ECOTOX - ECOTOXicology Knowledgebase System - ECOTOX Code Appendix', *Available at https://cfpub.epa.gov/ecotox/blackbox/help/codeappendix.pdf Accessed April 4, 2018* pp. 1–698.

US EPA (2017*c*), 'ECOTOX User Guide - ECOTOXicology Knowledgebase System Version 4.0', *Available at https://cfpub.epa.gov/ecotox/blackbox/help/userhelp4.pdf Accessed April 4, 2018* pp. 1–98.

US EPA (2017*d*), 'Pesticides Chemical Search | Chemical Search | Pesticides | US EPA', *Available at https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1 Accessed October-December 2017* .

US EPA (2017*e*), 'US EPA's Web Archive: Pesticide Reregistration Status', *Available at https://archive.epa.gov/pesticides/reregistration/web/html/status.html Accessed October-December 2017* .

Vanni, M. J. (2002), 'Nutrient cycling by animals in freshwater ecosystems', *Annual Review of Ecology and Systematics* **33**(1), 341–370.

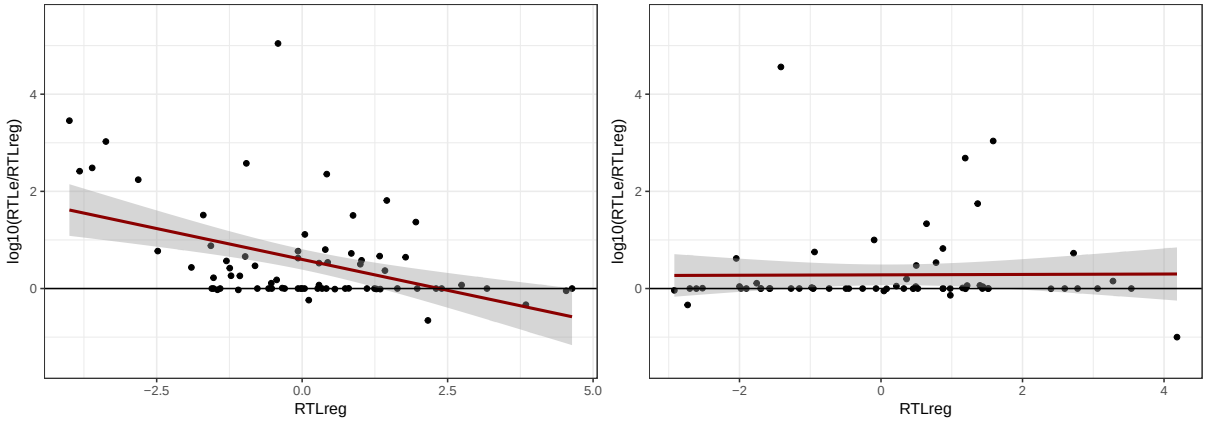
Wickham, H. (2009), 'ggplot2: Elegant Graphics for Data Analysis', *Available at http://ggplot2.org* .

WQP (2018), 'Water Quality Data Home', *Available at https://www.waterqualitydata.us/* .

A Appendix

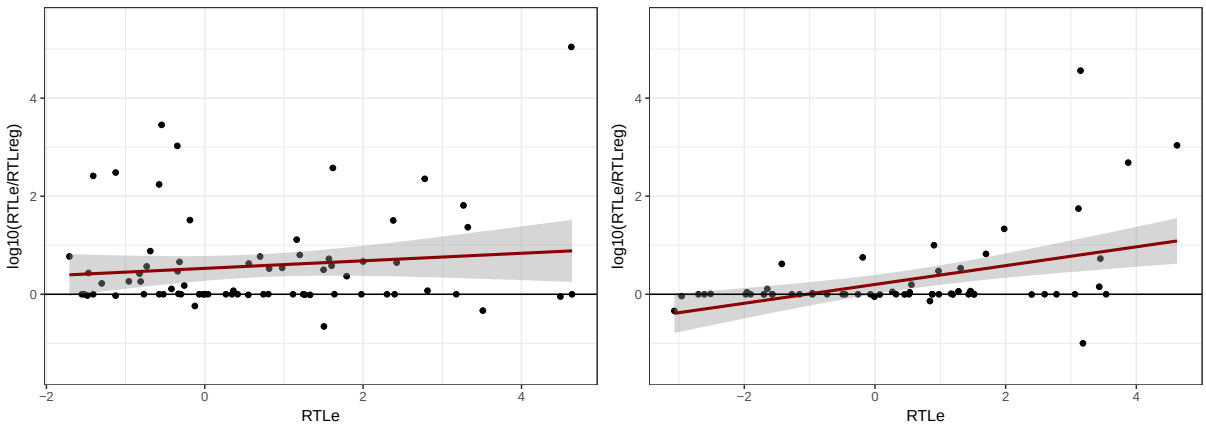
A.1 Supplemental Results

A.1.1 Visual inspection of correlations



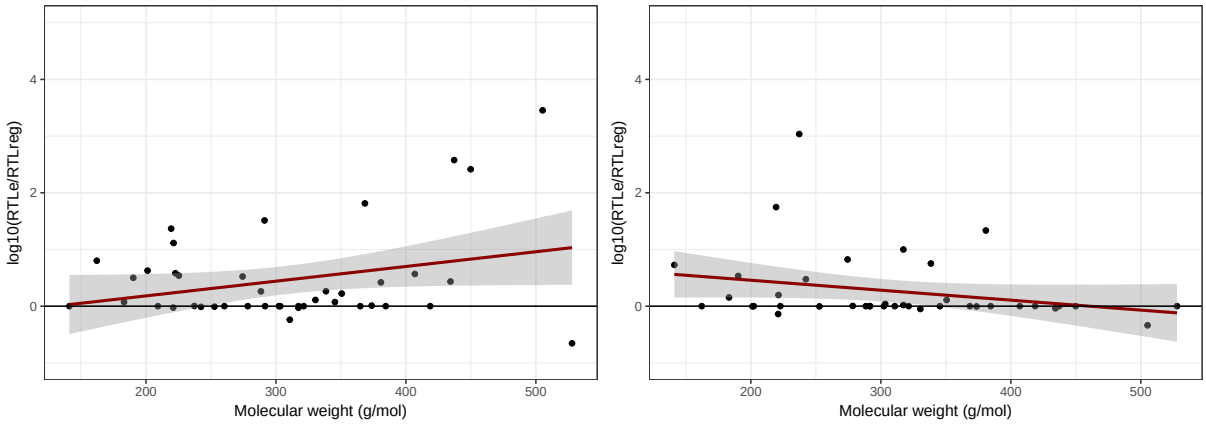
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A1: RTL_{reg} as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is linear relationship between the two variables (black dots) for freshwater model applications, but not for estuarine endpoints. Pearson's product-moment correlation *r* lay at -0.45 (95% CI -0.62 - (-0.25), *p*-value < 0.001, df = 70) for freshwater endpoints, and at 0.01 (95% CI -0.24 - 0.25, *p*-value = 0.947, df = 60) for estuarine endpoints.



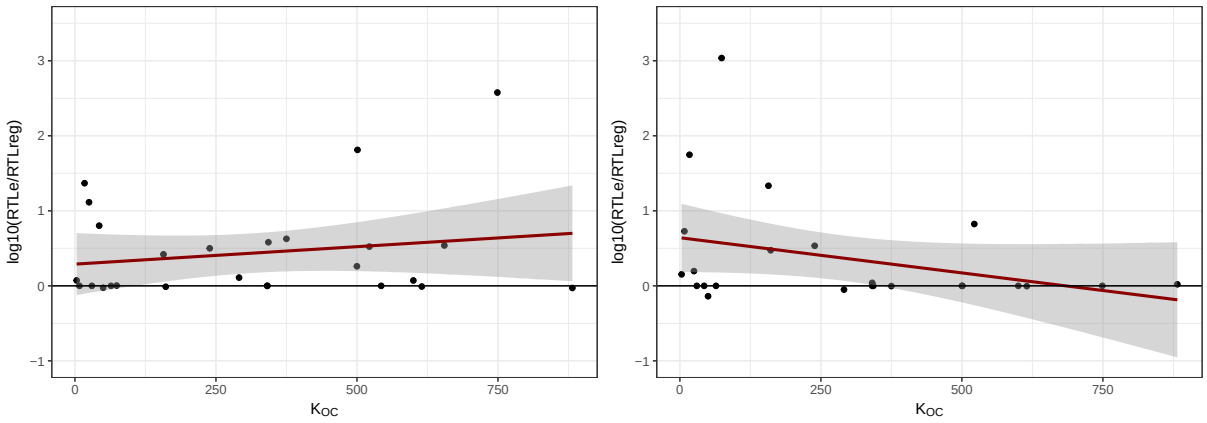
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A2: RTL_e as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is linear relationship between the two variables (black dots) for estuarine model applications, but not for freshwater endpoints. Pearson's product-moment correlation *r* lay at 0.12 (95% CI -0.11 - 0.34, *p*-value = 0.303, df = 70) for freshwater endpoints, and at 0.44 (95% CI 0.22 - 0.62, *p*-value < 0.001, df = 60) for estuarine endpoints.



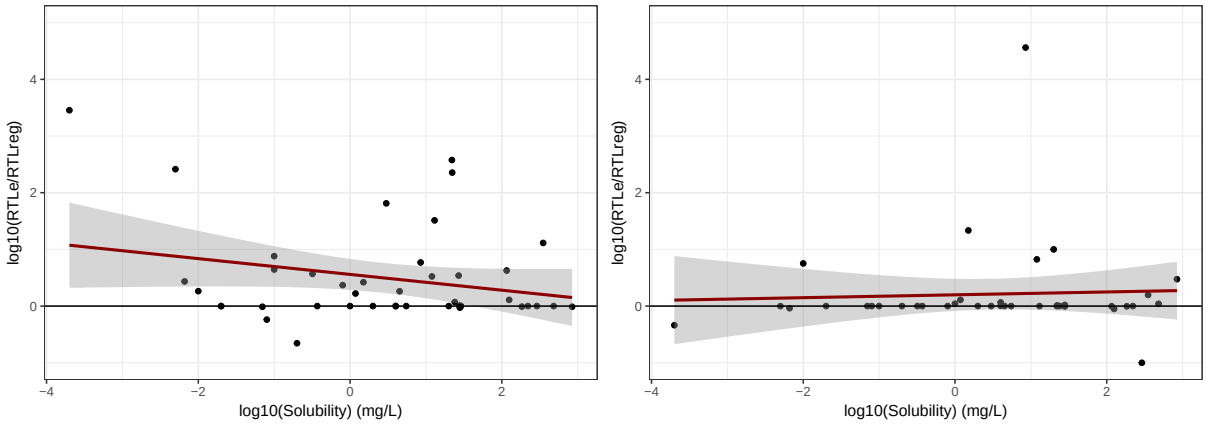
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A3: Molecular weight as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is linear relationship between the two variables (black dots) for freshwater model applications, but not for estuarine endpoints. Pearson's product-moment correlation *r* lay at 0.30 (95% CI 0.03 - 0.52, *p*-value = 0.003, df = 52) for freshwater endpoints, and at -0.17 (95% CI -0.43 - 0.12, *p*-value = 0.250, df = 47) for estuarine endpoints.



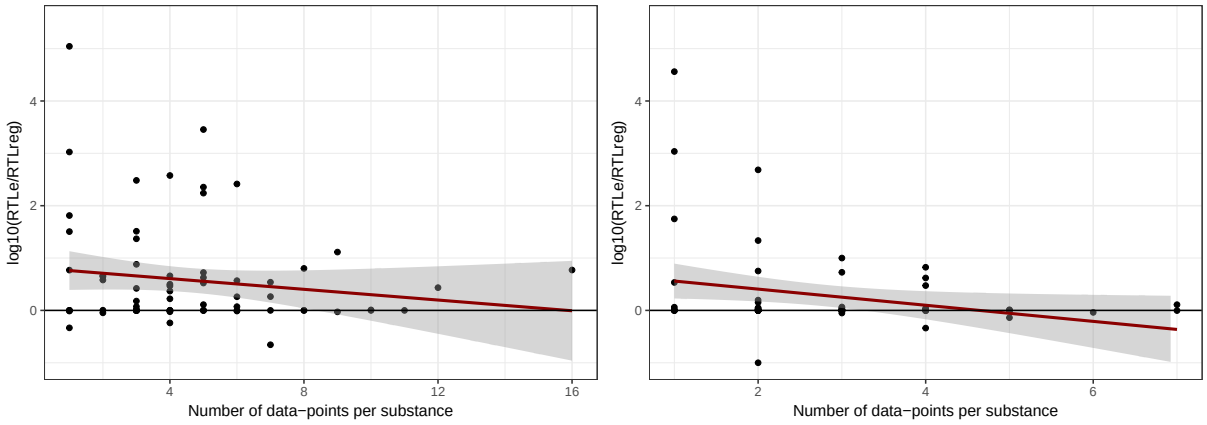
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A4: Organic carbon partitioning coefficient (K_{OC}) as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is no linear relationship between the two variables (black dots). Pearson's product-moment correlation *r* lay at 0.18 (95% CI -0.19 - 0.51, *p*-value = 0.332, df = 28) for freshwater endpoints, and at -0.35 (95% CI -0.64 - 0.04, *p*-value = 0.075, df = 25) for estuarine endpoints.



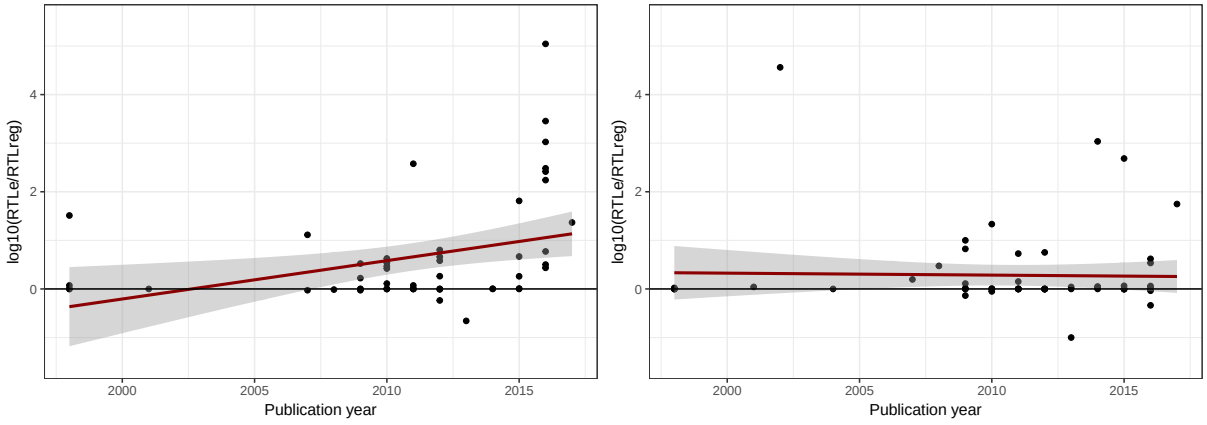
(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A5: Solubility as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is no linear relationship between the two variables (black dots). Pearson's product-moment correlation *r* lay at -0.25 (95% CI -0.52 - 0.06, *p*-value = 0.107, df = 40) for freshwater endpoints, and at 0.05 (95% CI -0.27 - 0.36, *p*-value = 0.772, df = 36) for estuarine endpoints.



(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A6: Number of datapoints per substance as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is no linear relationship between the two variables (black dots) for freshwater endpoints, but a significant relationship for estuarine endpoints. Pearson's product-moment correlation *r* lay at -0.15 (95% CI -0.37 - 0.09, *p*-value = 0.210, df = 70) for freshwater endpoints, and at -0.36 (95% CI -0.48 - 0.02, *p*-value = 0.038, df = 47) for estuarine endpoints.



(a) Freshwater endpoints **(b)** Estuarine endpoints

Figure A7: Publication year as a function of the ratio $\log_{10}(RTL_e/RTL_{reg})$ as a measure of prediction precision. Linear model smooth term (red line) with confidence bands (grey) indicates that there is linear relationship between the two variables (black dots) for freshwater model applications, but not for estuarine endpoints. Pearson's product-moment correlation *r* lay at 0.29 (95% CI 0.06 - 0.49, *p*-value = 0.014, df = 70) for freshwater endpoints, and at 0.03 (95% CI -0.27 - 0.23, *p*-value = 0.843, df = 60) for estuarine endpoints.

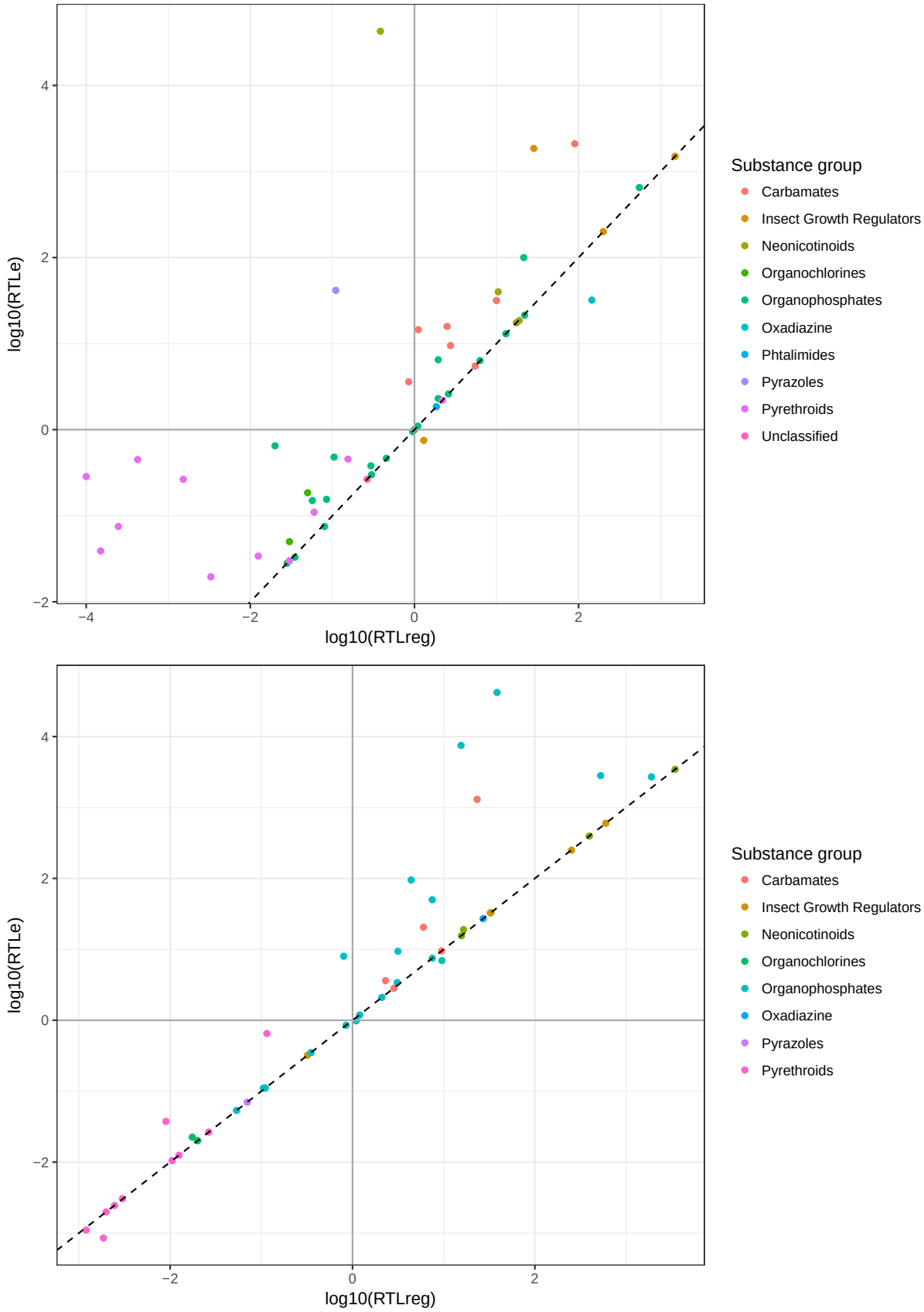


Figure A8: RTL_{reg} vs. RTL_e with substance groups (colors). If full model predictions were perfect, all points would lie on the dashed line. Predictions of data points above this line are to high, predictions of data points below the line are to low. There might be a problem with predicting Pyrethroids or very low RTL equivalents, since it appears as if they were likely to result in high overpredictions.

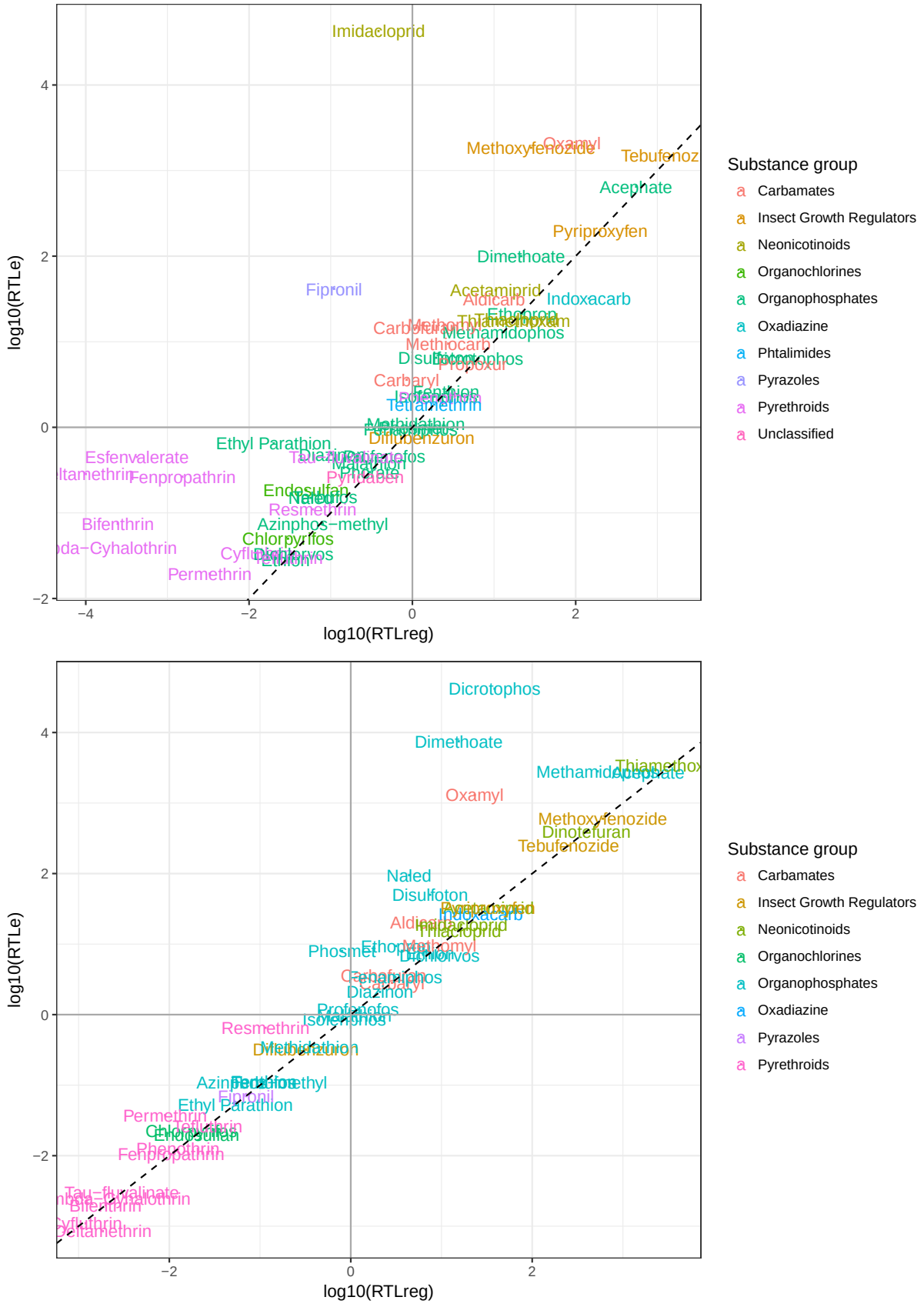


Figure A9: RTL_{reg} vs. RTL_e with insecticide names (text) and substance groups (colors). If full model predictions were perfect, all points would lie on the dashed line. Predictions of data points above this line are to high, predictions of data points below the line are to low. There might be a problem with predicting Pyrethroids or very low RTL equivalents, since it appears as if they were likely to result in high overpredictions.

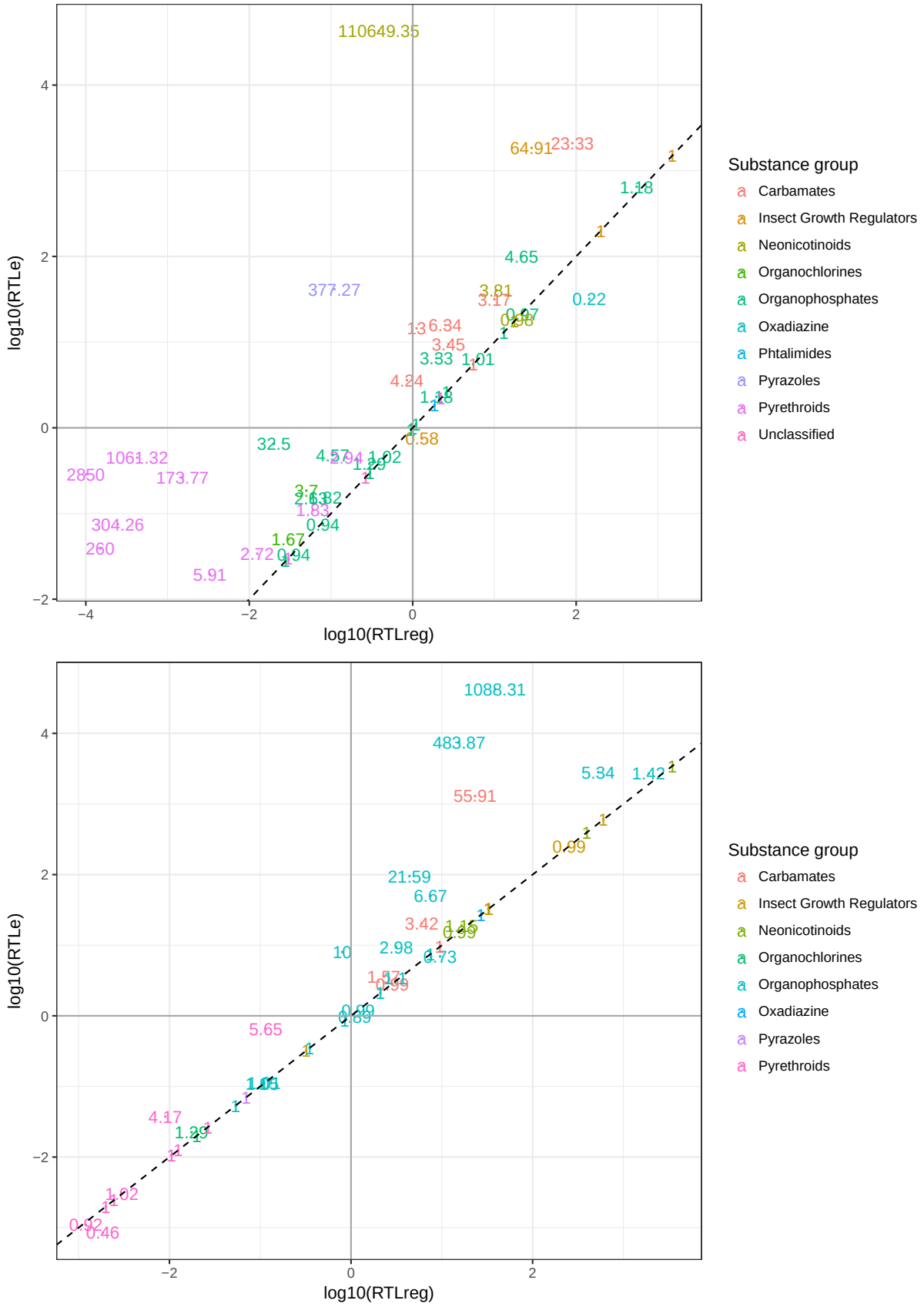


Figure A10: RTL_{reg} vs. RTL_e with ratios (RTL_e/RTL_{reg}) (numbers) and substance groups (colors). If full model predictions were perfect, all points would lie on the dashed line. Predictions of data points above this line are to high, predictions of data points below the line are to low. There might be a problem with predicting Pyrethroids or very low RTL equivalents, since it appears as if they were likely to result in high overpredictions.

A.1.2 Bootstrapped model precision

Table A1: Freshwater and estuarine model evaluation of prediction precision of mean proportion of datapoints which overestimates, meets, or underestimates the RTL_{reg} with 95% confidence levels for (a) null, mid and full model, and (b) the full model if data was included or not included in the ECOTOX database.

(a)	Freshwater Mean (95%-CI)	Estuaries Mean (95%-CI)	Model	Measure of precision
	0.927 (0.866 - 0.976)	0.901 (0.831 - 0.958)	null model	$\log_{10}(RTL_e/RTL_{reg}) < 0$
	0.024 (0.000 - 0.061)	0.056 (0.014 - 0.112)	null model	$\log_{10}(RTL_e/RTL_{reg}) = 0$
	0.049 (0.012 - 0.098)	0.042 (0.000 - 0.085)	null model	$\log_{10}(RTL_e/RTL_{reg}) > 0$
	0.333 (0.235 - 0.432)	0.217 (0.130 - 0.319)	mid model	$\log_{10}(RTL_e/RTL_{reg}) < 0$
	0.198 (0.123 - 0.296)	0.435 (0.319 - 0.551)	mid model	$\log_{10}(RTL_e/RTL_{reg}) = 0$
	0.469 (0.358 - 0.568)	0.348 (0.232 - 0.464)	mid model	$\log_{10}(RTL_e/RTL_{reg}) > 0$
	0.125 (0.056 - 0.208)	0.161 (0.081 - 0.258)	full model	$\log_{10}(RTL_e/RTL_{reg}) < 0$
	0.306 (0.208 - 0.417)	0.435 (0.323 - 0.548)	full model	$\log_{10}(RTL_e/RTL_{reg}) = 0$
	0.569 (0.458 - 0.681)	0.403 (0.290 - 0.532)	full model	$\log_{10}(RTL_e/RTL_{reg}) > 0$
(b)	**Freshwater Mean (95%-CI)**	**Estuaries Mean (95%-CI)**	**ECOTOX DB**	**Measure of precision**
	0.077 (0.000 - 0.192)	0.064 (0.000 - 0.016)	included	$\log_{10}(RTL_e/RTL_{reg}) < 0$
	0.846 (0.692 - 0.961)	0.871 (0.742 - 0.968)	included	$\log_{10}(RTL_e/RTL_{reg}) = 0$
	0.077 (0.000 - 0.192)	0.065 (0.000 - 0.161)	included	$\log_{10}(RTL_e/RTL_{reg}) > 0$
	0.152 (0.065 - 0.261)	0.258 (0.129 - 0.419)	not included	$\log_{10}(RTL_e/RTL_{reg}) < 0$
	0.848 (0.739 - 0.935)	0.742 (0.581 - 0.871)	not included	$\log_{10}(RTL_e/RTL_{reg}) > 0$

A.1.3 Model application

Table A2: Results of model application: Availability of RTL_es for set of insecticides from freshwater (FW) and estuarine (SW) model applications, for which no regulatory documents were available. Availability is indicated with an X, – indicates insecticides for which no RTL_es could be computed.

Substance	CAS number	Group	Authorized	RTL_e (FW|SW)
alpha-Endosulfan	959988	Organochlorines	No	– | –
alpha-Hexachlorocyclohexane	319846	Organochlorine	No	– | –
beta-Endosulfan	33213659	Organochlorines	No	– | –
delta-Hexachlorocyclohexane	319868	Organochlorine	No	– | –
(-)-cis-Permethrin	54774468	Pyrethroids	No	– | –
(-)-trans-Permethrin	54774479	Pyrethroids	No	– | –
Aldrin	309002	Organochlorine	No	– | –
Allethrin	584792	Pyrethroids	No	– | –
Bendiocarb	22781233	Carbamates	No	X | X
Carbophenothion	786196	Organophosphorus	No	– | –
Chlordane	57749	Organochlorines	No	– | –
Chlordane, technical	12789036	Organochlorines	No	– | –
cis-Chlordane	5103719	Organochlorines	No	– | –
Demeton	8065483	Organophosphates	No	X | –
Demeton-S	126750	Organophosphates	No	– | –

Table A2 continued: Results of model application: Availability of RTL_es for set of insecticides from freshwater (FW) and estuarine (SW) model applications, for which no regulatory documents were available. Availability is indicated with an X, – indicates insecticides for which no RTL_es could be computed.

Substance	CAS number	Group	Authorized	RTL_e (FW|SW)
Dieldrin	60571	Organochlorine	No	– | –
Endrin	72208	Organochlorine	No	– | –
Fenvalerate	51630581	Pyrethroids	No	X | X
Fonofos	944229	Organophosphorus	No	X | X
Mevinphos	7786347	Organophosphates	No	X | X
Mirex	2385855	Organochlorine	No	– | –
Novaluron	116714466	Insect Growth Regulators	No	– | X
o,p'-DDT	789026	Organochlorines	No	– | –
Omethoate	1113026	Organophosphates	No	– | –
Oxychlordane	27304138	Organochlorines	No	– | –
p,p'-DDT	50293	Organochlorines	No	– | –
Toxaphene	8001352	Organochlorine	No	– | –
trans-Chlordane	5103742	Organochlorines	No	– | –
trans-Nonachlor	39765805	Organochlorines	No	– | –

A.2 RTLE Model

A.2.1 Model input parameters

A.2.2 Regulatory threshold values

Table A3: Reviewed regulatory threshold values for the freshwater (fw) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants (non-vascular (non-vasc.) and vascular (vasc.))

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Abamectin	fw invert.	2013	71751412	Valid.	0.17
Acephate	fw invert.	2011	30560191	Calib.	550
Acetamiprid	fw invert.	2012	135410207	Calib.	10.5
Acetochlor	fw non-vasc.	2006	34256821	Herb.	1.43
Aldicarb	fw invert.	2016	116063	Calib.	10
Atrazin	fw vasc.	2016	1912249	Herb.	1
Azinphos-methyl	fw invert.	2007	86500	Calib.	0.08
Azoxystrobin	fw non-vasc.	2015	131860338	Fung.	49
Benomyl	fw fish	1997	17804352	Fung.	3.7
Bensulfuron-methyl	fw non-vasc.	2015	83055996	Herb.	7.8

Table A3 continued: Reviewed regulatory threshold values for the freshwater (fw) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants (non-vascular (non-vasc.) and vascular (vasc.))

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Bentazon	fw non-vasc.	2014	25057890	Herb.	4500
Beta Cypermethrin	fw fish	2016	65731842	Calib.	0.195
Bifenazate	fw invert.	2015	149877418	Valid.	250
Bifenthrin	fw invert.	2016	82657043	Calib.	0.0002465
Buprofezin	fw fish	2012	69327760	Valid.	165
Carbaryl	fw invert.	2010	63252	Calib.	0.85
Carbendazim	fw fish	2014	10605217	Fung.	5
Carbofuran	fw invert.	2007	1563662	Calib.	1.115
Chlorethoxyfos	fw invert.	2016	54593838	Valid.	0.027
Chlorimuron-ethyl	fw vasc.	2015	90982324	Herb.	0.27
Chlorothalonil	fw fish	2012	1897456	Fung.	9
Chlorpyrifos	fw invert.	2009	2921882	Calib.	0.03
Chlorsulfuron	fw vasc.	2015	64902723	Herb.	0.35
Clofentezine	fw fish	2014	74115245	Valid.	7.3
Cryolite	fw invert.	1996	15096523	Valid.	5000
Cyfluthrin	fw invert.	2016	68359375	Calib.	0.0125
Cyromazine	fw fish	2013	66215278	Valid.	42000
d-Allethrin	fw invert.	2014	584792	Calib.	1.05
Deltamethrin	fw invert.	2016	52918635	Calib.	1.00E-04
Diazinon	fw invert.	2012	333415	Calib.	0.105
Dichlorvos	fw invert.	2009	62737	Calib.	0.035
Dicofol	fw fish	2009	115322	Valid.	26.5
Dicrotophos	fw invert.	2014	141662	Calib.	6.3
Diflubenzuron	fw invert.	2012	35367385	Calib.	1.3
Dimethoate	fw invert.	2015	60515	Calib.	21.5
Dinotefuran	fw fish	2011	165252700	Calib.	49750
Disulfoton	fw invert.	2009	298044	Calib.	1.95
Diuron	fw non-vasc.	2009	330541	Herb.	2.4
Emamectin	fw invert.	2011	155569918	Valid.	0.5
Endosulfan	fw fish	2010	115297	Calib.	0.05
Esfenvalerate	fw invert.	2016	66230044	Calib.	0.000424
Ethion	fw invert.	1998	563122	Calib.	0.028
Ethoprop	fw invert.	2008	13194484	Calib.	22
Ethyl Parathion	fw invert.	1998	56382	Calib.	0.02

Table A3 continued: Reviewed regulatory threshold values for the freshwater (fw) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants (non-vascular (non-vasc.) and vascular (vasc.))

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Etoxazole	fw invert.	2014	153233911	Valid.	3.65
Fenamiphos	fw invert.	2001	22224926	Calib.	0.95
Fenbutatin Oxide	fw fish	2011	13356086	Valid.	0.85
Fenpropathrin	fw invert.	2016	39515418	Calib.	0.001525
Fenpyroximate	fw fish	2014	134098616	Valid.	0.365
Fenthion	fw invert.	1998	55389	Calib.	2.6
Fipronil	fw invert.	2011	120068373	Calib.	0.11
Flazasulfuron	fw vasc.	2015	104040780	Herb.	0.076
Flubendiamide	fw invert.	2010	272451657	Calib.	0.75
Flumetsulam	fw vasc.	2013	98967409	Herb.	3.1
Foramsulfuron	fw vasc.	2015	173159574	Herb.	0.56
Formetanate HCl	fw invert.	2004	23422539	Valid.	43.3
Halosulfuron-methyl	fw vasc.	2015	100784201	Herb.	0.042
Heptachlor	fw fish	1992	76448	Valid.	3.7
Hexythiazox	fw fish	2013	78587050	Valid.	60
Imidacloprid	fw invert.	2016	138261413	Calib.	0.385
Indoxacarb	fw fish	2013	173584446	Calib.	145
Iodosulfuron-methyl-Na	fw vasc.	2015	144550367	Herb.	0.7
Iprodione	fw non-vasc.	2012	36734197	Fung.	226
Isofenphos	fw invert.	1998	25311711	Calib.	1.95
Isoxaflutole	fw vasc.	2011	141112290	Fung.	4.9
Kresoxim-methyl	fw non-vasc.	2016	143390890	Fung.	30.3
lambda-Cyhalothrin	fw invert.	2016	91465086	Calib.	0.00015
Lindane	fw fish	2002	58899	Valid.	0.85
Malathion	fw invert.	2010	121755	Calib.	0.295
Mesosulfuron-methyl	fw vasc.	2015	400852666	Herb.	0.64
Metalaxyl	fw invert.	2010	57837191	Fung.	14000
Metaldehyde	fw fish	2006	108623	Valid.	34500
Metconazole	fw invert.	2005	125116236	Fung.	1050
Methamidophos	fw invert.	2011	10265926	Calib.	13
Methidathion	fw fish	2009	950378	Calib.	1.1
Methiocarb	fw invert.	2010	2032657	Calib.	2.75
Methomyl	fw invert.	2012	16752775	Calib.	2.5
Methoxychlor	fw invert.	1988	72435	Calib.	0.5

Table A3 continued: Reviewed regulatory threshold values for the freshwater (fw) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants (non-vascular (non-vasc.) and vascular (vasc.))

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Methoxyfenozide	fw invert.	2015	161050584	Calib.	28.5
Methyl parathion	fw invert.	2008	298000	Calib.	0.485
Metsulfuron-methyl	fw vasc.	2015	74223646	Herb.	0.36
Myclobutanil	fw non-vasc.	2009	88671890	Fung.	830
Naled	fw invert.	2010	300765	Calib.	0.057
Napropamide	fw vasc.	2005	15299997	Herb.	36
Orthosulfamuron	fw vasc.	2015	213464778	Herb.	0.7
Oryzalin	fw vasc.	2017	19044883	Herb.	13
Oxamyl	fw invert.	2017	23135220	Calib.	90
Oxydemeton-Methyl	fw invert.	2008	301122	Valid.	95
Pentachlorophenol	fw fish	2014	87865	Valid.	7.5
Permethrin	fw invert.	2016	52645531	Calib.	0.0033
Phenothrin	fw invert.	2012	26002802	Calib.	2.2
Phorate	fw invert.	2009	298022	Calib.	0.3
Phosmet	fw invert.	2010	732116	Calib.	1
Primisulfuron-methyl	fw vasc.	2015	86209510	Herb.	0.8
Profenofos	fw invert.	2015	41198087	Calib.	0.455
Propargite	fw invert.	2014	2312358	Valid.	7
Propiconazole	fw non-vasc.	2012	60207901	Fung.	21
Propoxur	fw invert.	2014	114261	Calib.	5.5
Prosulfuron	fw vasc.	2015	94125345	Herb.	1.22
Pymetrozine	fw invert.	2013	123312890	Valid.	43500
Pyraclostrobin	fw non-vasc.	2014	175013180	Fung.	1.5
Pyridaben	fw invert.	2010	96489713	Calib.	0.265
Pyriproxyfen	fw invert.	2011	95737681	Calib.	200
Resmethrin	fw invert.	2012	10453868	Calib.	0.06
Spinosad	fw invert.	2016	131929607	Valid.	7000
Sulfometuron-methyl	fw non-vasc.	2015	74222972	Herb.	4.3
Sulfosulfuron	fw vasc.	2015	74223566	Herb.	0.985
Tau-fluvalinate	fw invert.	2010	102851069	Calib.	0.155
Tebuconazole	fw vasc.	2009	107534963	Fung.	151.5
Tebufenozide	fw fish	2015	112410238	Calib.	1500
Tefluthrin	fw fish	2012	79538322	Calib.	0.03
Terbacil	fw non-vasc.	1998	5902512	Herb.	11

Table A3 continued: Reviewed regulatory threshold values for the freshwater (fw) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants (non-vascular (non-vasc.) and vascular (vasc.))

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Terbufos	fw invert.	2015	13071799	Calib.	0.085
Tetraconazole	fw vasc.	2005	112281773	Fung.	310
Tetramethrin	fw fish	2010	7696120	Calib.	1.85
Thiacloprid	fw invert.	2012	111988499	Calib.	18.9
Thiamethoxam	fw invert.	2011	153719234	Calib.	17.5
Thifensulfuron-methyl	fw vasc.	2015	79277273	Herb.	1.59
Thiodicarb	fw invert.	2009	59669260	Valid.	2.65
Triasulfuron	fw vasc.	2015	82097505	Herb.	0.19
Tribenuron-methyl	fw vasc.	2015	101200480	Herb.	2
Tribufos	fw invert.	2008	78488	Herb.	3.4
Trifloxystrobin	fw non-vasc.	2017	141517217	Fung.	0.974
Trifloxysulfuron-Na	fw vasc.	2015	199119589	Herb.	0.24
Tebupirimphos	fw invert.	2009	96182535	Valid.	0.039

Table A4: Reviewed regulatory threshold values for the estuarine (est.) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants.

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Abamectin	est. invert.	2013	71751412	Valid.	0.01
Acephate	est. invert.	2011	30560191	Calib.	1900
Acetamiprid	est. invert.	2012	135410207	Calib.	33
Acetochlor	est. invert.	2006	34256821	Herb.	1100
Aldicarb	est. invert.	2016	116063	Calib.	6
Atrazin	est_plant	2016	1912249	Herb.	12
Azinphos-methyl	est. invert.	1998	86500	Calib.	0.105
Azoxystrobin	est. invert.	2015	131860338	Fung.	28
Benomyl	est. invert.	1997	17804352	Fung.	49
Bifenazate	est. invert.	2015	149877418	Valid.	25
Bifenthrin	est. invert.	2016	82657043	Calib.	0.001985
Buprofezin	est. invert.	2012	69327760	Valid.	93.5
Carbaryl	est. invert.	2010	63252	Calib.	2.85
Carbofuran	est. invert.	2007	1563662	Calib.	2.3
Chlorethoxyfos	est. invert.	2016	54593838	Valid.	0.027
Chlorothalonil	est. invert.	2012	1897456	Fung.	1.8

Table A4 continued: Reviewed regulatory threshold values for the estuarine (est.) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants.

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Chlorpyrifos	est. invert.	2009	2921882	Calib.	0.0175
Cyfluthrin	est. invert.	2016	68359375	Calib.	0.0012
Cyromazine	est. invert.	2013	66215278	Valid.	52000
Deltamethrin	est. invert.	2016	52918635	Calib.	0.00185
Diazinon	est. invert.	2012	333415	Calib.	2.1
Dichlorvos	est. invert.	2009	62737	Calib.	9.55
Dicofol	est. invert.	1998	115322	Valid.	7.55
Dicrotophos	est. invert.	2014	141662	Calib.	38.5
Diflubenzuron	est. invert.	2012	35367385	Calib.	0.32
Dimethoate	est. invert.	2015	60515	Calib.	15.5
Dinotefuran	est. invert.	2011	165252700	Calib.	395
Disulfoton	est. invert.	2009	298044	Calib.	7.5
Diuron	est. invert.	2003	330541	Herb.	500
Emamectin	est. invert.	2011	155569918	Valid.	0.02
Endosulfan	est. invert.	2010	115297	Calib.	0.02
Esfenvalerate	est. invert.	2016	66230044	Calib.	0.00233
Ethion	est. invert.	1998	563122	Calib.	7.5
Ethoprop	est. fish	2008	13194484	Calib.	3.15
Ethyl Parathion	est. invert.	1998	56382	Calib.	0.0535
Etoxazole	est. invert.	2014	153233911	Valid.	0.55
Fenamiphos	est. invert.	2001	22224926	Calib.	3.1
Fenbutatin Oxide	est. invert.	2011	13356086	Valid.	0.185
Fenpropathrin	est. invert.	2016	39515418	Calib.	0.0105
Fenpyroximate	est. invert.	2014	134098616	Valid.	1.65
Fenthion	est. invert.	1998	55389	Calib.	0.11
Fipronil	est. invert.	2011	120068373	Calib.	0.07
Flubendiamide	est. invert.	2010	272451657	Calib.	14
Formetanate HCl	est. invert.	2004	23422539	Valid.	1150
Halosulfuron-methyl	est. invert.	2015	100784201	Herb.	47000
Hexythiazox	est. invert.	2013	78587050	Valid.	55
Imidacloprid	est. invert.	2016	138261413	Calib.	16.5
Indoxacarb	est. invert.	2013	173584446	Calib.	27.1
Iprodione	est. invert.	2012	36734197	Fung.	1150
Isofenphos	est. invert.	1998	25311711	Calib.	0.85
Isoxaflutole	est. invert.	2011	141112290	Fung.	8.9

Table A4 continued: Reviewed regulatory threshold values for the estuarine (est.) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants.

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Kresoxim-methyl	est. invert.	2016	143390890	Fung.	7.5
lambda-Cyhalothrin	est. invert.	2016	91465086	Calib.	0.002455
Lindane	est. invert.	2002	58899	Valid.	0.0385
Malathion	est. invert.	2010	121755	Calib.	1.1
Metalaxyl	est. invert.	2010	57837191	Fung.	12850
Methamidophos	est. invert.	2011	10265926	Calib.	527
Methidathion	est. invert.	2009	950378	Calib.	0.35
Methomyl	est. invert.	2012	16752775	Calib.	9.5
Methoxychlor	est. invert.	1988	72435	Calib.	1.8
Methoxyfenozide	est. invert.	2015	161050584	Calib.	600
Methyl parathion	est. invert.	2006	298000	Calib.	0.175
Naled	est. fish	2010	300765	Calib.	4.4
Napropamide	est. invert.	2005	15299997	Herb.	700
Oryzalin	est. invert.	2017	19044883	Herb.	14.25
Oxamyl	est. invert.	2017	23135220	Calib.	23.25
Pentachlorophenol	est. fish	2014	87865	Calib.	15.5
Permethrin	est. invert.	2016	52645531	Calib.	0.009
Phenothrin	est. invert.	2012	26002802	Calib.	0.0125
Phorate	est. invert.	2009	298022	Calib.	0.055
Phosmet	est. invert.	2009	732116	Calib.	0.8
Profenofos	est. invert.	2015	41198087	Calib.	1.2
Propargite	est. fish	2014	2312358	Valid.	27.5
Propiconazole	est. invert.	2012	60207901	Fung.	56.5
Propoxur	est. invert.	2014	114261	Calib.	20.5
Pymetrozine	est. invert.	2013	123312890	Valid.	15250
Pyraclostrobin	est. invert.	2014	175013180	Fung.	2.08
Pyriproxyfen	est. invert.	2011	95737681	Calib.	32.5
Resmethrin	est. invert.	2012	10453868	Calib.	0.115
Spinosad	est. invert.	2016	131929607	Valid.	150
Tau-fluvalinate	est. invert.	2010	102851069	Calib.	0.003
Tebuconazole	est. invert.	2009	107534963	Fung.	245
Tebufenozide	est. invert.	2015	112410238	Calib.	252.5
Tefluthrin	est. invert.	2012	79538322	Calib.	0.0265
Terbacil	est. invert.	1998	5902512	Herb.	2450
Terbufos	est. invert.	2015	13071799	Calib.	0.11

Table A4 continued: Reviewed regulatory threshold values for the estuarine (est.) model calibration (Calib.), validation (Valid.) of insecticides and herbicides (Herb.) and fungicides (Herb). Most sensitive endpoints are reported. Endpoint types were invertebrates (invert.), fish, and aquatic plants.

Substance	Endpoint type	Year	CAS number	Priority	RTL_{reg} (ug/L)
Tetraconazole	est. invert.	2005	112281773	Fung.	220
Thiacloprid	est. invert.	2012	111988499	Calib.	15.65
Thiamethoxam	est. invert.	2011	153719234	Calib.	3450
Thiodicarb	est. invert.	2009	59669260	Valid.	14.25
Tribufos	est. invert.	1996	78488	Herb.	2.5
Trifloxystrobin	est. invert.	2017	141517217	Fung.	4.31
Tebupirimphos	est. invert.	2009	96182535	Valid.	3.315

A.2.3 List of reviewed registration documents:

1. EPA, Addendum to Data Evaluation Report on the 10-Day Sediment Toxicity Test with the Freshwater Amphipod (Hyalella azteca). Washington, DC. 2015
2. EPA, Azardirachtin Summary Document Registration Review: Initial Docket September 2008. Washington, DC. 2008
3. EPA, Azinphos-methyl, Case Number: 0234, Chemical Number, 058001. 1999. IN: Memorandum EFED RED Chapter for azinphos methyl, Case Number: 0234, DP Bar Code D234006, D234029, D234030. Washington, DC. 1998
4. EPA, Ecological Effects Branch Science Chapter for benomyl Reregistration Eligibility Document (RED). Washington, DC. 1997
5. EPA, ecological risk assessment for Fipronil Uses. Washington, DC. 2007
6. EPA, ecological risk assessment in Support of Registration Review of Clofentezine. Washington, DC. 2014
7. EPA, EFED Problem Formulation for Registration Review for Ethoprofl (Ethoprophos). Washington, DC. 2008
8. EPA, EFED RED Chapter for Ethyl Parathion. Washington, DC. 1998
9. EPA, EFED RED Chapter for Isofenphos. Washington, DC. 1998
10. EPA, EFED RED Chapter Tribufos. Washington, DC. 1996
11. EPA, EFED Registration Review Preliminary Problem Formulation for Resmethrin. Washington, DC. 2012
12. EPA, EFED Registration Review Preliminary Problem Formulation for Thiacloprid Washington, DC. 2012
13. EPA, EFED Registration Review Problem Formulation for Abamectin. Washington, DC. 2013

14. EPA, EFED Registration Review Problem Formulation for Fenpyroximate. Washington, DC. 2014

15. EPA, EFED Reregistration Review: Preliminary Problem Formulation for Isocaflutole. Washington, DC. 2011

16. EPA, EFED Science Chapter for the Formetanate Hydrochloride Reregistration Eligibility Document. Washington, DC. 2004

17. EPA, Endosulfan:2010 Environmental Fate and ecological risk assessment. Washington, DC. 2010

18. EPA, Environmental Fate and ecological risk assessment for the Proposed New Use of d-Phenothrin in Outdoor Rediential Misting Systems. Washington, DC. 2012

19. EPA, Environmental Fate and Effects Division Risk Assessment for Legume Vegetable and Christmas tree New Uses for Insecticide Flubendiamide. Washington, DC. 2010

20. EPA, EPA, Registration Review: Preliminary Problem Formulation for the Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Pyraclostrobin. Washington, DC. 2014

21. EPA, Evaluation of Potential Ecological Risks from the Proposed New Use of Tebuconazole on Fruiting Vegetables (Crop Group 8). Washington, DC. 2009

22. EPA, Fenamiphos Environmental Risk Assessment by EFED's Fenamiphos RED Team. Washington, DC. 2001

23. EPA, Kresoxim-methyl – Transmittal of the Preliminary Environmental Fate and ecological risk assessment for Registration Review. Washington, DC. 2016

24. EPA, Level I Screening ecological risk assessment for the Reregistration of Metaldehyde. Washington, DC. 2006

25. EPA, Methoxychlor Registration Standard. Washington, DC. 1988

26. EPA, Oxamyl: Preliminary ecological risk assessment for Registration Review. Washington, DC. 2017

27. EPA, Pesticide Fact Sheet Tetraconazole. Waschington, DC. 2005

28. EPA, Preliminary Aquatic Risk Assessment to Support the Registration Review of Imidacloprid. Washington, DC. 2016

29. EPA, Preliminary Comparative Environmental Fate and ecological risk assessment for the Registration Review of Eight Synthetic Pyrethroids and the Pyrethrins. Washington, DC. 2016

30. EPA, Preliminary ecological risk assessment for Registration Review of 22 Sulfonylurea Herbicides. Washington, DC. 2015

31. EPA, Preliminary ecological risk assessment for the Registration Review of Chlorethoxyfos. Washington, DC. 2016

32. EPA, Preliminary ecological risk assessment for the Registration Review of Cyromazine. Washington, DC. 2013

33. EPA, Preliminary ecological risk assessment for the Registration Review of Flumetsulam. Washington, DC. 2013

34. EPA, Preliminary ecological risk assessment for the Registration Review of Trifloxystrobin. Washington, DC. 2017

35. EPA, Preliminary Environmental Fate and ecological risk assessment for the Registration Review of Spinosad. Washington, DC. 2016

36. EPA, Preliminary Environmental Fate and ecological risk assessment for the Registration Review of Propoxur, Washington, DC. 2014

37. EPA, Preliminary Problem Formulation for the Ecological Risk and Drinking Water Exposure Assessments for the Registration Review of Buprofezin. Washington, DC. 2012

38. EPA, Preliminary Problem Formulation for the Environmental Fate, Ecological Risk, Endangered Species, and Human Health Drinking Water Exposure Assessments in Support of the Registration Review of Thiophanate Methyl and Carbendazim. Washington, DC. 2014

39. EPA, Problem Formulation for the Ecological Risk and Drinking Water Exposure Assessments in Support of the Registration Review of Dinotefuran. Washington, DC. 2011

40. EPA, Problem Formulation for the Environmental Fate and Ecological Risk, and Endangered Species Assessments in Support of the Registration Review of Metalaxyl and Mefenoxam (Metalaxyl-M). Washington, DC. 2010

41. EPA, Problem Formulation for the Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Assessments in Support of the Registration Review of Fenbutatinoxide. Washington, DC. 2011

42. EPA, Proposed Interim Registration Review Decision for Aldicarb (Case# 0140, PC Code 098301). Washington, DC. 2016

43. EPA, Registration Review - Preliminary Problem Formulation for Ecological Risk and Environmental Fate, Endangered Species and Drinking Water Assessments for Prosulfuron. Washington, DC. 2012

44. EPA, Registration Review - Preliminary ecological risk assessment for Dimethoate. Washington, DC. 2015

45. EPA, Registration Review - Preliminary Problem Formulation for the ecological risk assessment of Dichlorvos (DDVP). Washington, DC. 2009”

46. EPA, Registration Review - Preliminary Problem Formulation for the Ecological Risk and Drinkin Water Exposure Assessments for Tau-Fluvalinate. Washington, DC. 2010

47. EPA, Registration Review - Preliminary Problem Formulation for the Ecological Risk and Drinking Water Exposure Assessments for Methidathion. Washington, DC. 2009

48. EPA, Registration Review - Preliminary Problem Formulation for the ecological risk assessment of Disulfoton. Washington, DC. 2009

49. EPA, Registration Review - Preliminary Problem Formulation for the ecological risk assessment of Oxydemeton-Methyl. Washington, DC. 2008

50. EPA, Registration Review - Preliminary Problem Formulation for the Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Emamectin Benzoate. Washington, DC. 2011

51. EPA, Registration Review - Revised Preliminary Problem Formulation for the ecological risk assessment of Phorate. Washington, DC. 2009

52. EPA, Risks of Phorate Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*) to Federally Threatened Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*) to Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*) to Federally Endangered San Joaquin Kit Fox (*Vulpes macrotis mutica*). Washington, DC. 2008

53. EPA, Registration Review – Preliminary Problem Formulation for Ecological Risk and Environmental Fate, Endangered Species, and Drinking Water Assessments for Fipronil. Washington, DC. 2011

54. EPA, Registration Review – Preliminary Problem Formulation for Ecological Risk and Environmental Fate, Endangered Species, and Drinking Water Assessments for Iprodione. Washington, DC. 2012

55. EPA, Registration Review – Preliminary Problem Formulation for the Ecological Risk and Drinking Water Exposure Assessments for Tebupirimphos. Washington, DC. 2009

56. EPA, Registration Review ecological risk assessment and Effects Determination for Sodium Bentazon. Washington, DC. 2014

57. EPA, Registration Review ecological risk assessment for Bifenazate. Washington, DC. 2015

58. EPA, Registration Review Problem Formulation for Indoxacarb. Washington, DC. 2013

59. EPA, Registration Review Problem Formulation for Pyriproxyfen. Washington, DC. 2011

60. EPA, Registration Review- Preliminary Problem Formulation for the ecological risk assessment and Drinking Water Exposure Assessment to Be Conducted for Etoxazole. Washington, DC. 2014

61. EPA, Registration Review: Draft ecological risk assessment and Endangered Species Effects Determination for d-Allethrin. Washington, DC. 2014

62. EPA, Registration Review: Draft Risk Assessment of the Environmental Fate and Ecological Risk of Azoxystrobin. Washington, DC. 2015

63. EPA, Registration Review: ecological risk assessment for Terbufos. Washington, DC. 2015

64. EPA, Registration Review: Preliminary Environmental Fate and ecological risk assessment Endangered Species Effects Determination for Methoxyfenozide. Washington, DC. 2015

65. EPA, Registration Review: Preliminary Environmental Fate and ecological risk assessment for Oryzalin. Washington, DC. 2017

66. EPA, Registration Review: Preliminary Problem Formulation for Environmental Fate and Ecological Risk, Endangered Species, and Drinking Water Assessments for Phosmet. Washington, DC. 2009

67. EPA, Registration Review: Preliminary Problem Formulation for Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Methiocarb. Washington, DC. 2010

68. EPA, Registration Review: Preliminary Problem Formulation for Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for pyridaben. Washington, DC. 2010

69. EPA, Registration Review: Preliminary Problem Formulation for Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Tefluthrin. Washington, DC. 2012

70. EPA, Registration Review: Preliminary Problem Formulation for Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Thiodicarb. Washington, DC. 2009

71. EPA, Registration Review: Preliminary Problem Formulation for Pymetrozine. Washington, DC. 2013

72. EPA, Registration Review: Problem Formulation for the Environmental Fate, Ecological Risk, Endangered Species, and Drinking Water Exposure Assessments for Thiamethoxam. Washington, DC. 2011

73. EPA, Registration Review: Problem Formulation for the Environmental Fate, Ecological Risk, ndangered Species, and Drinking Water Exposure Assessments for Diflubenzuron. Washington, DC. 2012

74. EPA, Registratrion Review - Preliminary Problem Formulation for Ecological Risk and Environmental Fate, Endangered Species and Drinking Water Assessments for Acetamiprid. Washington, DC. 2012

75. EPA, Reregistration Eligibility Decision (RED) Cryolite. Washington, DC. 1996

76. EPA, Reregistration Eligibility Decision (RED) Dicofol. Washington, DC. 1998

77. EPA, Reregistration Eligibility Decision (RED) Document for Tetramethrin. Washington, DC. 2010

78. EPA, Reregistration Eligibility Decision (RED) TERBACIL. Washington, DC. 1998

79. EPA, REREGISTRATION ELIGIBILITY DECISION Fenthion EFED CHAPTER. Washington, DC. 1998

80. EPA, Reregistration Eligibility Decision for Diuron. Washington, DC.

81. EPA, Reregistration Eligibility Decision for Methyl Parathion. Washington, DC. 2006

82. EPA, Reregistration Eligibility Decision for Napropamide. Washington, DC. 2005

83. EPA, Reregistration Eligibility Document Heptachlor. Washington, DC. 1992

84. EPA, Revised EFED RED Chapter for Lindane, Washington DC. 2002

85. EPA, Risks of Acephate Use to the Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*), Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*), and California Tiger Salamander (*Ambystoma californiense*), Central California Distinct Population Segment And the Federally Endangered California Clapper Rail (*Rallus longirostris obsoletus*), California Freshwater Shrimp (*Syncaris pacifica*), California Tiger Salamander (*Ambystoma californiense*) Sonoma County Distinct Population Segment and Santa Barbara County Distinct Population Segment, Salt Marsh Harvest Mouse (*Reithrodontomys raviventris*), San Francisco Garter Snake (*Thamnophis sirtalis tetrataenia*), and San Joaquin Kit Fox (*Vulpes macrotis mutica*). Washington, DC. 2011

86. EPA, Risks of Azinphos Methyl Use to the Federally Listed California Red Legged Frog. Washington, DC. 2007

87. EPA, Risks of Carbaryl Use to Federally Threatened Delta Smelt (*Hypomesus transpacificus*). Washington, DC. 2010

88. EPA, Risks of Chlorothalonil Use to Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*), California Tiger Salamander (*Ambystoma californiense*), Central California Distinct Population Segment, and Delta Smelt (*Hypomesus transpacificus*), And the Federally Endangered California Clapper Rail (*Rallus longirostris obsoletus*), California Freshwater Shrimp (*Syncaris pacifica*), California Tiger Salamander (*Ambystoma californiense*) Sonoma County Distinct Population Segment and Santa Barbara County Distinct Population Segment, San Francisco Garter Snake (*Thamnophis sirtalis tetrataenia*), and Tidewater Goby (*Eucyclogobius newberryi*). Washington, DC. 2012

89. EPA, Risks of Chlorpyrifos Use to Federally Threatened & Endangered California red-legged frog (*Rana aurora draytonii*), California tiger salamander (*Ambystoma californiense*), San Francisco garter snake (*Thamnophis sirtalis tetrataenia*), California clapper rail, (*Rallus longirostris obsoletus*), Salt marsh harvest mouse (*Reithrodontomys raviventris*), Bay checkerspot butterfly (*Euphydryas editha bayensis*), Valley elderberry longhorn beetle (*Desmocerus californicus dimorphus*), San Joaquin kit fox (*Vulpes macrotis mutica*), California freshwater shrimp (*Syncaris pacifica*), and Delta smelt (*Hypomesus transpacificus*). Washington, DC. 2009

90. EPA, Risks of Cyfluthrin and Beta-Cyfluthrin Use To Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*), Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*), California Tiger Salamander (*Ambystoma californiense*), Central California Distinct Population Segment, and Delta Smelt (*Hypomesus transpacificus*), And the Federally Endangered California Clapper Rail (*Rallus longirostris obsoletus*), California Freshwater Shrimp (*Syncaris*

pacificus), California Tiger Salamander (*Ambystoma californiense*) Sonoma County Distinct Population Segment and Santa Barbara County Distinct Population Segment, San Francisco Garter Snake (*Thamnophis sirtalis tetrataenia*), and Tidewater Goby (*Eucyclogobius newberryi*). Washington, DC. 2013

91. EPA, Risks of Diazinon Use to Federally Threatened Delta Smelt (*Hypomesus transpacificus*) and the Federally Endangered Tidewater Goby (*Eucyclogobius newberryi*). Washington, DC. 2012

92. EPA, Risks of Dicofol Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*). Washington, DC. 2009

93. EPA, Risks of Diuron Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*). Washington, DC. 2009

94. EPA, Risks of Malathion Use to the Federally Threatened Delta Smelt (*Hypomesus transpacificus*) and California Tiger Salamander (*Ambystoma californiense*), Central California Distinct Population Segment, and the Federally Endangered California Tiger Salamander, Santa Barbara County and Sonoma County Distinct Population Segments. Washington, DC. 2010

95. EPA, Risks of Methidathion Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*) Pesticide Effects. Washington, DC. 2009

96. EPA, Risks of Methomyl Use to Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*), Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*), California Tiger Salamander (*Ambystoma californiense*), Central California Distinct Population Segment, and Delta Smelt (*Hypomesus transpacificus*), And the Federally Endangered California Clapper Rail (*Rallus longirostris obsoletus*), California Freshwater Shrimp (*Syncaris pacificus*), California Tiger Salamander (*Ambystoma californiense*) Sonoma County Distinct Population Segment and Santa Barbara County Distinct Population Segment, San Francisco Garter Snake (*Thamnophis sirtalis tetrataenia*), and Tidewater Goby (*Eucyclogobius newberryi*). Washington, DC. 2012

97. EPA, Risks of Methyl Parathion Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*). Washington, DC. 2008

98. EPA, Risks of Myclobutanil Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*). Washington, DC. 2009

99. EPA, Risks of Naled Use to Federally Threatened Bay Checkerspot Butterfly (*Euphydryas editha bayensis*), and Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*), And the Federally Endangered California Clapper Rail (*Rallus longirostris obsoletus*), San Francisco Garter Snake (*Thamnophis sirtalis tetrataenia*), and San Joaquin Kit Fox (*Vulpes macrotis mutica*). Washington, DC. 2010; EPA, Risks of Naled Use to Federally Threatened ... Appendix M. Washington, DC. 2010

100. EPA, Risks of Phosmet Use to the Federally Threatened and Endangered California Tiger Salamander (*Ambystoma californiense*). Washington, DC. 2010

101. EPA, Risks of Tribufos Use to Federally Threatened California Red-legged Frog (*Rana aurora draytonii*). Washington, DC. 2008

102. EPA, Section 18 ecological risk assessment for the Control of Asian Rust on Soybeans using Metconazole. Washington, DC. 2005

103. EPA, Section 3 Environmental Risk Assessment for the New Use Registration of Acetochlor on Sorghum and Sweet Corn. Washington, DC. 2006

104. EPA, The Risk Assessment and Science Support Branch/Antimicrobials Division's Science Chapter for the Reregistration Eligibility Decision Document (RED) for Pentachlorophenol. Washington, DC. 1999

105. EPA, Transmittal of EFED List A Summary Report for Ethion. Washington, DC. 1998

106. EPA, Transmittal of Preliminary Environmental Fate and ecological risk assessment for Registration Review of the Insect Growth Regulator, Tebufenozide. Washington, DC. 2015

107. EPA, Transmittal of the Draft Environmental Fate and ecological risk assessment in Support of the Registration Review of Profenofos. Washington, DC. 2015

108. EPA, Transmittal of the Preliminary Environmental Fate and ecological risk assessment for the Registration of Dicrotophos. Washington, DC. 2014

109. EPA, Transmittal of the Preliminary Environmental Fate and ecological risk assessment in Support of the Registration Review of Hexythiazox. Washington, DC. 2013

110. EPA, Updated ecological risk assessment for the Proposed New Use of Propiconazole on sugarcane. Washington, DC. 2012

111. EPA, US Environmental Protection Agency Office of Pesticide Programs Reregistration Eligibility Decision for Carbofuran. Washington, DC. 2007

A.2.4 Model species

Table A5: List of standard test species with source as selection reason. Group indicates whether species are plants (P) or animals (A). Asterisks (*) indicate old species names.

Species	Group	Source
Aedes taeniorhynchus	A	Reported endpoints
Americamysis bahia	A	OCSPP 850.1035: Mysid Acute Toxicity Test
Anabaena flos-aquae	P	OCSPP 850.4550: Cyanobacteria (*Anabaena flos-aquae*) Toxicity
C. dubia	A	Reported endpoints
Ceriodaphnia dubia	A	Reported endpoints
Chironomus dilutus	A	OCSPP 850.1735: Spiked Whole Sediment 10-Day Toxicity Test, Freshwater Invertebrates
Chironomus pulmosus	A	Reported endpoints

Table A5 continued: List of standard test species with source as selection reason. Group indicates whether species are plants (P) or animals (A). Asterisks (*) indicate old species names.

Species	Group	Source
Chironomus riparius	A	OECD 235
Chironomus sp.	A	OECD 235
*Chironomus tentans**	A	OCSPP 850.1735: Spiked Whole Sediment 10-Day Toxicity Test, Freshwater Invertebrates
Chironomus yoshimatsui	A	OECD 218
Chloroperla grammatica	A	Expert judgement
Cloeon dipterum	A	Reported endpoints
Crassostrea gigas	A	OCSPP 850.1055: Bivalve Acute Toxicity Test (Embryo-Larval)
Crassostrea virginica	A	OCSPP 850.1025: Oyster Acute Toxicity Test (Shell Deposition)
Cyprinodon variegatus	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Cyprinus carpio	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Danio rerio	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Daphnia magna	A	OCSPP 850.1010: Aquatic Invertebrate Acute Toxicity Test, Freshwater Daphnids
Daphnia pulex	A	OCSPP 850.1010: Aquatic Invertebrate Acute Toxicity Test, Freshwater Daphnids
Daphnia sp.	A	OECD 202
Esox lucius	A	Reported endpoints
Farfantepenaeus aztecus	A	OCSPP 850.1045: Penaeid Acute Toxicity Test
Farfantepenaeus duorarum	A	OCSPP 850.1045: Penaeid Acute Toxicity Test
Fundulus similes	A	Reported endpoints
Gammarus fasciatus	A	OCSPP 850.1020: Gammarid Amphipod Acute Toxicity Test
Gammarus lacustris	A	OCSPP 850.1020: Gammarid Amphipod Acute Toxicity Test
Gammarus pseudolimnaeus	A	OCSPP 850.1020: Gammarid Amphipod Acute Toxicity Test
Hyalella azteca	A	OCSPP 850.1735: Spiked Whole Sediment 10-Day Toxicity Test, Freshwater Invertebrates
Ictalurus punctatus	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Isochrysis galbana	A	Reported endpoints

Table A5 continued: List of standard test species with source as selection reason. Group indicates whether species are plants (P) or animals (A). Asterisks (*) indicate old species names.

Species	Group	Source
Isoperla grammatica	A	Reported endpoints
Isoperla sp.	A	Expert judgement
Jappa kutera	A	Reported endpoints
Lagodon rhomboides	A	Reported endpoints
Leiostomus xanthurus	A	Reported endpoints
Lemna gibba	P	OCSPP 850.4400: Aquatic Plant Toxicity Test Using *Lemna spp.*
Lepomis macrochirus	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Leuciscus idus	A	Reported endpoints
Litopenaeus setiferus	A	OCSPP 850.1045: Penaeid Acute Toxicity Test
Marone saxatilis	A	Reported endpoints
Menidia beryllina	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Menidia menidia	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Menidia peninsulae	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Menippe mercenaria	A	Reported endpoints
Mercenaria mercenaria	A	OCSPP 850.1055: Bivalve Acute Toxicity Test (Embryo-Larval)
*Mysidopsis bahia**	A	OCSPP 850.1035: Mysid Acute Toxicity Test
Mytilus edulis	A	OCSPP 850.1055: Bivalve Acute Toxicity Test (Embryo-Larval)
Navicula	A	Reported endpoints
Navicula pelliculosa	A	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Oncorhynchus kisutch	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Oncorhynchus mykiss	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Orconectes nais	A	Reported endpoints
Oryzias latipes	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Palaemonetes pugio	A	Reported endpoints
*Penaeus aztecus**	A	OCSPP 850.1045: Penaeid Acute Toxicity Test
*Penaeus duorarum**	A	OCSPP 850.1045: Penaeid Acute Toxicity Test

Table A5 continued: List of standard test species with source as selection reason. Group indicates whether species are plants (P) or animals (A). Asterisks (*) indicate old species names.

Species	Group	Source
Penaeus onorarum	A	Reported endpoints
*Penaeus setiferus**	A	OCSPP 850.1045: Penaeid Acute Toxicity Test
Perca flavescens	A	Reported endpoints
Pimephales promelas	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Poecilia reticulata	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Pseudokirchneriella subcapitata	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Pteronarcella badia	A	Expert judgement
Pteronarcella sp.	A	Expert judgement
Pteronarcys	A	Reported endpoints
*Raphidocelis subcapitata**	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Pteronarcys californica	A	Reported endpoints
Salmo salar	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Salvelinus fontinalis	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Salvelinus namaycush	A	Reported endpoints
Selenastrum capricornutum	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Simocephalus serrulatus	A	Reported endpoints
Simulium vittatum	A	Reported endpoints
Skeletonema costatum	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Thalassoma bifasciatum	A	OCSPP 850.1075: Freshwater and Saltwater Fish Acute Toxicity Test
Lemna minor	P	OCSPP 850.4400: Aquatic Plant Toxicity Test Using *Lemna spp.*
Desmodesmus subspicatus	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
*Scenedesmus subspicatus**	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Microcystis aeruginosa	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity

Table A5 continued: List of standard test species with source as selection reason. Group indicates whether species are plants (P) or animals (A). Asterisks (*) indicate old species names.

Species	Group	Source
Thalassiosira pseudonana	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Dunaliella tertiolecta	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity
Phaeodactylum tricornutum	P	Ecological Effects Test Guidelines OCSPP 850.4500: Algal Toxicity

A.2.5 SQL code for model application

```
SELECT DISTINCT ON (test_cas) test_cas, rtle0
            FROM (
        ---- Calculate RTLe from endpoints for animals (LOC acute 0.5),
        ---- and aquatic plants (LOC acute 1)
            SELECT
            tests.test_cas,
            CASE
            WHEN
            species.kingdom = 'Animalia'
            THEN
            results.conc1_mean::numeric * unit_convert.multiplier::numeric * .5
            ELSE
            results.conc1_mean::numeric * unit_convert.multiplier::numeric
            END rtle0
            FROM
            ecotox_updated.results
        ---- Join tables
            INNER JOIN
            ecotox_updated.unit_convert ON results.conc1_unit = unit_convert.unit
            INNER JOIN
            ecotox_updated.tests ON results.test_id = tests.test_id
            INNER JOIN
            ecotox_updated.species ON tests.species_number = species.species_number
            INNER JOIN
            ecotox_updated.model_species ON species.latin_name = model_species.species
            INNER JOIN
            ecotox_updated.refs ON tests.reference_number = refs.reference_number
         -- INNER JOIN
         -- ecotox_updated.model_cas_fw ON tests.test_cas = model_cas_fw.test_cas
            --! add join for validation and calibration of model to access CAS numbers and
                regulatory threshold values for model substances
            --! ESTUARINE MODEL: replace by
                -- ecotox_updated.model_cas_est ON tests.test_cas = model_cas_est.test_cas

        ---- Apply filter criteria
--- Null model: ---
          --# filter 0:
            WHERE results.conc1_mean::numeric > 0
            AND test_cas IN(959988, 33213659, 22781233)
                -- paste respective cas numbers
            AND unit_convert.unit_conv = 'ug/L'
            AND tests.media_type = 'FW'
```

```
            --!  ESTUARINE MODEL: change to 'SW'
            --!  AQUATIC PLANT MODEL: replace by:
                -- AND (tests.media_type = 'FW'
                --     OR (species.kingdom IN ('Plantae', 'Chromista')
                --     AND tests.media_type IN ('SW', 'NC', 'NR', 'UKN')))

---  Mid model:     ---
        --# filter 1:
            AND species.kingdom NOT IN ('Plantae', 'Chromista')
            --! remove for AQUATIC PLANT MODEL
        --# filter 2:
            AND results.effect IN ('MOR', 'ITX')
        --# filter 3:
            AND tests.test_type = 'ACUTE'
        --# filter 4:
            AND results.endpoint ~ '^(L|E|I)C50'
        --# filter 5:
            AND results.obs_duration_mean || results.obs_duration_unit IN ('48h', '72h', '96h',
                '2d', '3d', '4d')
            --!  AQUATIC PLANT MODEL: replace by:
                -- AND (results.obs_duration_mean || results.obs_duration_unit IN ('48h', '72h', '96
                    h', '2d', '3d', '4d')
                --    OR species.kingdom IN ('Plantae', 'Chromista')
                --    AND results.obs_duration_mean || results.obs_duration_unit IN ('120h', '5d', '7
                    d', '14d'))

---  Full model:     ---
        --# filter 6:
            AND results.measurement !~ '^MULT'
        --# filter 7:
            AND tests.other_effect_comments = ''
        --# filter 8:
            AND (tests.test_purity_mean::numeric >= 70 OR tests.test_purity_mean IN ('NC', 'NR
                '))
        --# filter 9:
            AND results.conc1_type IN ('A', 'NC', 'NR')
        --# filter 10:
            AND tests.test_location = 'LAB'
        --# filter 11
        -- AND model_cas_fw.year::bigint >= refs.publication_year::bigint
            --! time filter , for calibration and validation only
        --# filter 12:
            AND (results.conc1_min NOT IN ('NR') AND results.conc1_max NOT IN ('NR'))
        --# filter 13:
            AND tests.organism_source NOT IN ('MLT', 'WLD')
        --# filter 14:
            AND results.conc1_mean_op = ''
        --# filter 15:
            AND tests.test_method NOT IN ('NC', 'NR')
        --# filter 16:
            AND results.chem_analysis_method != 'U'
        --# filter 17:
            AND tests.control_type NOT IN ('Z', 'U', 'I', 'O', 'H', '␣')
        --# filter 18:
            AND tests.organism_characteristics NOT IN ('>24␣hr','>24␣Hr', 'Fry')
)   considered_endpoints
            ORDER BY test_cas, rtle0;
```

A.3 Validity elements

A.3.1 Summary of standard guideline requirements

Table A6: Test Elements for the OCSPP Acute Toxicity Test Guideline Studies for aquatic animals (US EPA, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)

Criteria	Acute Daphnia, OCSPP 850.1010	Acute Gammarids, OCSPP 850.1020	Acute Oyster, OCSPP 850.1025	Acute Mysid, OCSPP 850.1035	Acute Peneid, OCSPP 850.1045	Acute Bivalve, OCSPP 850.1055	Acute FW/SW Fish, OCSPP 850.1075
Test type	Static, static-renewal, or flow-through	Static, static-renewal, or flow-through	Flow-through	Static, static-renewal, or flow-through	Static, static-renewal or flow-through	Static	Static, static-renewal, or flow-through
Test species	Daphnia magna or Daphnia pulex	Recommended *Gammarus fasciatus*, *Gammarus pseudolimnaeus*, or *Gammarus lacustris*	*Crassostrea virginica*	*Americamysis bahia*	*Farfantepenaeus aztecus* (brown shrimp); *Farfantepenaeus duorarum* (pink shrimp); or *Litopenaeus setiferus* (white shrimp)	*Crassostrea virginica* (Eastern oysters) is the preferred test species; other test species that may be used include *Crassostrea gigasn* (Pacific oysters), *Mercenaria mercenaria* (quahogs, hard clams), or *Mytilus edulis* (blue mussels, bay mussels)	Freshwater species. *Salmo salar*; *Lepomis macrochirus*; *Salvelinus fontinalis*; *Ictalurus punctatus*; *Oncorhynchus kisutch*; *Cyprinus carpio*; *Pimephales promelas*; *Poecilia reticulata*; *Oncorhynchus mykiss*; *Oryzias latipes*; *Danio rerio*. Saltwater species. *Menidia menidia*; *Menidia beryllina*; *Menidia peninsulae*; *Cyprinodon variegatus*;
Test duration	48 hours	96 hours	96 hours	96 hours	96 hours	48 hours	96 hours
Temperature	Selected from a range of 18 to 22 °C (preferably 20 °C) (constant during test within ±1 °C of selected test temperature)	18 °C (constant during test within ±1 °C)	20 °C (constant during test within ±2 °C), However, if unfiltered natural seawater that has not been previously held is used, temporary fluctuations (less than 8 hours) of ±5 °C may occur and be tolerated by oysters (i.e., not affect control performance) due to their adaptations to fluctuating tidal habitats.	25 °C (constant during test within ±1 °C)	23 °C (constant during test within ±1 °C)	Varies by species (see Table 1) (constant during test within ±1 °C of selected test temperature)	12 °C for Atlantic salmon, Brook trout, Coho salmon, Rainbow trout 22 °C for Atlantic silverside, Bluegill sunfish, Channel catfish, Common carp, Sheepshead minnow, Inland silverside, Tidewater silverside, Fathead minnow, Guppy, Medaka, Zebrafish, Constant during test within ±2 °C
Light quality	Ambient laboratory illumination	Ambient laboratory illumination	Ambient laboratory illumination	Ambient laboratory illumination	Ambient laboratory illumination	Ambient laboratory illumination	Ambient laboratory illumination
Light intensity	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)	540-1080 lux (approximately 50-100 ft-c)

Table A6 continued: Test Elements for the OCSPP Acute Toxicity Test Guideline Studies for aquatic animals (US EPA, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)

Criteria	Acute Daphnia, OCSPP 850.1010	Acute Gammarids, OCSPP 850.1020	Acute Oyster, OCSPP 850.1025	Acute Mysid, OCSPP 850.1035	Acute Peneid, OCSPP 850.1045	Acute Bivalve, OCSPP 850.1055	Acute FW/SW Fish, OCSPP 850.1075
Photoperiod	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark schemes	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark schemes	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark schemes	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark with a recommended 30-minute transition period	Prefer 12 hours light:12 hours dark with a recommended 30-minute transition period	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark schemes	Selected from among 12 hours light:12 hours dark to 16 hours light:8 hours dark schemes
pH	Between 6.0 and 8.5 (constant during test within ±1 pH unit)	Between 6.0 and 8.5 (constant during test within ±1 pH unit)	Between 7.5 and 8.5 (constant during test within ±1 pH unit)	Between 7.5 and 8.5 (constant during test within ±1 pH unit)	Between 7.5 and 8.5 (constant during test within ±1 pH unit)	Between 7.5 and 8.5 (constant during test within ±1 pH unit)	Between 6.0 and 8.5 for freshwater testing; between 7.5 and 8.5 for saltwater testing (constant during test within ±1 pH unit)
Water hardness (as CaCO3)	<250 mg/L (preferably <180 mg/L); for testing with metals, 40-50 mg/L	<250 mg/L (preferably <180 mg/L); for testing with metals, 40-50 mg/L	Artificial or natural seawater that is diluted with freshwater: 20 ppt (range of ±2 ppt during test); Natural seawater that is not diluted with freshwater to reduce salinity: >12 ppt (range of <5 ppt during test)	20 ppt (constant during test within ±2 ppt)	20 ppt (constant during test within ±2 ppt)	Artificial or natural seawater that is diluted with freshwater: 20 ppt (range of ±2 ppt during test); Natural seawater that is not diluted with freshwater to reduce salinity: >12 ppt (range of <5 ppt during test)	For freshwater: <250 mg/L (preferably <180 mg/L); 40-50 mg/L for testing with metals
Salinity	–	–	–	–	–	–	Selected from a range of 15 to 25 ppt (constant during test within ±2 ppt or selected salinity)
TOC	≤2 mg/L	<2 mg/L	≤2 mg/L	Less than or equal to (≤) 2 mg/L	Less than or equal to (≤) 2 mg/L	≤2 mg/L	≤2 mg/L
Age of test organisms	<24 hours old	<24 hours post-release or early instar (first or second)	30-50 mm in valve height; similar in age and size; standard deviation of valve height <20% of the mean	<24-hour post-release mysids	Post-larval juveniles	Embryos at 2- to 8-cell stage at test initiation	Juvenile fish <3.0 grams
Number of test organisms per concentration	20 (minimum)	20 (minimum)	Multiple-concentration definitive test: 20 (minimum), Limit test: 32 (minimum)	20 (minimum)	20 (minimum)	Multiple-concentration definitive test: 2 (minimum; more are preferable for the control(s)), Limit test: 4 (minimum)	7 minimum, 10 preferred

Table A6 continued: Test Elements for the OCSPP Acute Toxicity Test Guideline Studies for aquatic animals (US EPA, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)

Criteria	Acute Daphnia, OCSPP 850.1010	Acute Gammarids, OCSPP 850.1020	Acute Oyster, OCSPP 850.1025	Acute Mysid, OCSPP 850.1035	Acute Peneid, OCSPP 850.1045	Acute Bivalve, OCSPP 850.1055	Acute FW/SW Fish, OCSPP 850.1075
Number of replicate test vessels per concentration	2 (minimum)	2 (minimum)	Multiple-concentration definitive test: 2 (minimum), Limit test: 4 (minimum)	2 (minimum)	2 (minimum)	15-30 embryos/mL	1 minimum, 2 preferred
Loading	Static or static-renewal tests: ≥20 mL per daphnid; Flow-through test: ≤0.5 g/L per 24 hours and <5 g/L at any time	Static or static-renewal tests: ≤0.8 g/L;	Flow rate is adequate to promote adequate shell growth and maintain environmental conditions. Flow rate of 1 L/h/individual has been shown to be adequate when using unfiltered natural seawater that is not supplemented with additional algae	Static or static-renewal tests: No more than 30 mysids/L; Flow-through test: ≤0.5 g/L per 24 hours and <5 g/L at any time	Static or static-renewal tests: ≤0.8 g/L; Flow-through test: ≤0.5 g/L per 24 hours and <5 g/L at any time	≤30 embryos/mL	Static or static-renewal tests: ≤0.8 g wet weight per liter; Flow-through test: ≤0.5 g/L per 24 hours and <5 g/L at any time
Feeding regime	No feeding during test	No feeding during test	Phytoplankton naturally occurring in the dilution water (if using unfiltered, unsterilized natural seawater) or supplemented (needed if using artificial seawater)	Daily, with a live food such as *Artemia spp. nauplii*	No feeding during test	No feeding during test	No feeding during test
Test vessel aeration	Strongly discouraged because a single air bubble can become caught under the carapace of a daphnid and kill it or float the daphnid to the surface where it will become trapped. Gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Not recommended; gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Not recommended; gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Not recommended; gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Not recommended; gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Strongly discouraged because bubbles can collect within the mantle cavity of a larva resulting in death. Gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.	Not recommended; gentle aeration of test vessels may only be used in cases where the dissolved oxygen levels are in danger of dropping below 60% saturation. In such cases, assurances should be made that the use of aeration does not stress the test organisms; test substance concentrations should be measured during the test; and all treatment and control vessels should be given the same aeration treatment.

Table A6 continued: Test Elements for the OCSPP Acute Toxicity Test Guideline Studies for aquatic animals (US EPA, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)

Criteria	Acute Daphnia, OCSPP 850.1010	Acute Gammarids, OCSPP 850.1020	Acute Oyster, OCSPP 850.1025	Acute Mysid, OCSPP 850.1035	Acute Peneid, OCSPP 850.1045	Acute Bivalve, OCSPP 850.1055	Acute FW/SW Fish, OCSPP 850.1075
Test concentrations	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used	Definitive test: minimum of 5 test concentrations chosen in a geometric series plus a dilution water control and a vehicle (solvent) control, if a vehicle is used
Vehicle concentration, if used	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)	≤0.1 mL/L for recommended solvents (see OCSPP 850.1000)
Measure of effect or measurement endpoint	48-h EC50 based on immobilization	96-h LC_{50} based on mortality	96-h IC_{50} based on reduction in shell growth	96-h LC_{50} based on mortality	96-h LC_{50} based on mortality	48-h EC50 based on the percentage of larvae that are dead or failed to develop complete shells	96-h LC_{50} based on mortality

Table A7: Test Elements for the OCSPP Acute Toxicity Test Guideline Studies for aquatic plants (US EPA, 2012*a*,*b*,*c*)

Criteria	**OCSPP 850.4500: Algal Toxicity**	**OCSPP 850.4400: Aquatic Plant Toxicity Test Using *Lemna spp.***	**OCSPP 850.4550: Cyanobacteria (*Anabaena flos-aquae*) Toxicity**
Test type	Static	Static renewal (pesticides); Static, static renewal, or flow through (industrial chemicals)	Static
Test duration	96 hours	7 days	96 hours
Test matrix	Synthetic growth medium appropriate for the test species	Synthetic growth medium	Synthetic growth medium
Temperature	24 °C for *P. subcapitata* and *N. pelliculosa*; 20 °C for *S. costatum*. Constant during test within ± 2 °C.	25 ± 2 °C	24 ± 2 °C
Light quality	Cool-white fluorescent	Warm-white or cool-white fluorescent	Cool-white fluorescent
Light intensity	60 µmol/m2/s	57 - 90 µmol/m2/s	30 µmol/m2/s
Photoperiod	Continuous light for *P. subcapitata* and *N. pelliculosa*; 14 hour light:10 hour dark for *S. costatum*	Continuous	Continuous light
Shaking	Continuous at 100 oscillations per minute for *P. subcapitata* and *N. pelliculosa*; manual, once or twice daily, for *S. costatum*	–	Manual, once or twice daily
Salinity	30 ± 5 ppt for saltwater species (*S. costatum*)	–	–
Test vessel size	125 - 500 mL Erlenmeyer flasks	Sufficient to prevent crowding (e.g., 250 - 1000 mL beakers or flasks)	125 - 500 mL Erlenmeyer flasks
Test solution volume	Less than or equal to 50% of the volume of the test vessel	From healthy stock cultures 7 -12 days old	Less than or equal to 50% of the volume of the test vessel
Age of inoculum	From logarithmically-growing stock cultures (typically 3 - 7 days old)	12 -16 fronds total, with the same number of plants and fronds in each test vessel	From logarithmically- growing stock cultures (typically 3 - 7 days old)
Inoculum concentration	10,000 cells per milliliter (cells/mL) for *P. subcapitata* and *S. costatum*. At least 10,000 cells/mL for other species. Inoculum volume less than 2 mL.	–	10,000 cells/mL. Inoculum volume less than 2 mL.
Number of replicate test vessels per concentration	Four (minimum)	Four (minimum)	Four (minimum)
Test concentrations	Unless performing limit test, minimum of 5 test concentrations plus appropriate controls	Unless performing limit test, minimum of 5 test concentrations plus appropriate controls	Unless performing limit test, minimum of 5 test concentrations plus appropriate controls
Test concentration preparation	Aqueous solutions prepared by adding test substance to synthetic nutrient medium, directly or via vehicle	Aqueous solutions prepared by adding test substance to synthetic nutrient medium, directly or via vehicle	Aqueous solutions prepared by adding test substance to synthetic nutrient medium, directly or via vehicle
Measures of effect (Measurement endpoints)	96-hour IC_{50} and NOEC (or IC_{05}) values for yield, average specific growth rate, and area under the growth curve based on algal cell density	IC_{50} and NOEC (or IC_{05}) values for yield, and average specific growth rate based on frond number IC_{50} and NOEC (or IC_{05}) values for yield and average specific growth rate based on frond size (dry weight or frond area)	96-hour IC_{50} and NOEC (or IC_{05}) values for yield, average specific growth rate, and area under the growth curve based on algal cell density

A.3.2 Validity criteria

Table A8: Validity Criteria for the OCSPP Acute Toxicity Test Guideline Studies for aquatic organisms (US EPA, 2012*a*,*b*,*c*, 2016*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*)

	Standard test validity criteria	Freshwater Daphnids	Gammarid Amphipods	Oyster	*A. bahia*	Penaeid	Bivalve	Fish	Algal	Lemna	Cyanobacteria
1.	All test vessels (and retention chambers for *A. bahia*, closures for plants) were not identical.	x	x	x	x	x	x	x	x	x	x
2.	Treatments were not randomly or indiscriminately assigned to individual test vessel locations, or individual test organisms were not randomly or indiscriminately assigned to test vessels (and retention chambers for *A. bahia*, positions in the growth chamber for plants).	x	x	x	x	x	x	x	x	x	x
3.	A medium/dilution water control (and vehicle (solvent) control, if a vehicle was used) was not included in the test.	x	x	x	x	x	x	x	x	x	x
4.	More than 10% of the organisms in either the dilution water or vehicle (solvent) controls showed signs of disease, stress (e.g., discoloration, unusual behavior, immobilization, shell gaping, excessive mucus), and/or death.	x	x	x	x	x	–	x	–	–	–
5.	Test organisms were fed during the test.	x	x	–	–	–	–	x	–	–	–
6.	A surfactant or dispersant was used in the preparation of a stock or test solution. (However, adjuvants may be used when testing pesticide typical end-use products.)	x	x	x	x	x	x	x	x	x	x
7.	An overall mean of at least 2 mm of new shell growth (i.e., overall mean of all replicates) was not observed in each control group (vehicle (solvent) control and dilution water control).	–	–	x	–	–	–	–	–	–	–
8.	Evidence of spawning was observed.	–	–	x	–	–	–	–	–	–	–
9.	Less than 70% of oyster embryos or 60% of hard clam or mussel embryos in either the dilution water control or vehicle (solvent) control resulted in normal larvae at test termination.	–	–	–	–	–	x	–	–	–	–
10.	The concentration of solvent in the range used affected growth of the test species.	–	–	–	–	–	–	–	x	x	x
11.	During the 96 hour test period, cell counts in the controls did not increase by a factor of at least 100 times for *P. subcapitata* and a factor of at least 30 times for *S. costatum* (i.e., logarithmic growth in the controls was not reached during the test).	–	–	–	–	–	–	–	x	–	–
12.	A minimum of five test concentrations were not used in the definitive test.	–	–	–	–	–	–	–	x	x	x
13.	Controls were contaminated with the test substance.	–	–	–	–	–	–	–	x	x	x
14.	The lowest test concentration level was not less than the 96-hour yield, average specific growth rate, and area under the growth curve IC_{50} values based on cell density.	–	–	–	–	–	–	–	x	–	x
15.	Temperature and light intensity were not measured as specified during the test.	–	–	–	–	–	–	–	x	x	x
16.	The doubling time of number of fronds in the control exceed 2.5 days.	–	–	–	–	–	–	–	–	x	–
17.	The lowest test concentration level was not less than the 7-day yield and average specific growth rate IC_{50} values based on number of fronds and frond area (dry weight or frond area).	–	–	–	–	–	–	–	–	x	–

A.3.3 Rejection reasons

Table A9: Rejection Rate Analysis (US EPA, 1994)

	Rejection reason	Bluegill (Freshwater)	Rainbow (Freshwater)	Invertebrate (Freshwater)	Fish (Estuarine)	Mollusks (Estuarine)	Shrimps (Estuarine)
1a.	The concentration level selected was less than 100 mg/L and was not high enough to produce an LC_{50}.	x	x	–	–	–	–
1b.	The concentration levels selected were less than 100 mg/L and because of solubility problems were not high enough to produce an LC_{50}	x	x	–	x	–	–
1c.	There were insufficient concentration levels to result in percent mortality to mollusk embryos/larvae of greater than 65%, thus resulting in a statistically less-reliable LC_{50}.	–	–	–	–	x	–
1d.	Solubility needed to achieve LC_{50} was not obtained (and was questionable, since previously submitted data appeared to indicate that high levels of solubility could be achieved)	–	–	–	–	–	x
2.	The test chambers were aerated.	x	x	–	–	–	–
3.	The biological loading of test vessels was twice the recommended amount.	x	x	–	–	–	–
4.	Test substance purity was not identified.	x	x	–	–	–	x
5.	Inappropriate test species were used and test species were not clearly identified.	x	x	–	–	–	–
6.	Fish were fed during 96-hour exposure period of the study.	x	x	–	–	–	–
7.	Minimum limit detectability, or the minimum quantifiable limit, was not defined quantitatively.	x	x	x	–	–	–
8.	The variability limit for test concentrations was greater than 1.5.	x	x	x	–	–	–
9.	The dilution water contained higher levels than recommended of lead, iron, and aluminum.	x	x	x	–	–	–
10a.	There was no solvent control.	x	x	–	–	–	–
10b.	The type of solvent used in the study, and the amount used in the solvent control and each treatment level, were not provided.	–	–	–	–	–	x
11.	The results of several of the test concentrations were obtained from a separate test conducted a few weeks after the definitive study.	x	x	–	–	–	–
12a.	Not all test solutions were measured 96 hours as well as at zero hours. The concentration of the test material greatly decreased for at least one test level after 96 hours of exposure.	x	–	–	–	–	–
12b.	Not all test solutions were measured 96 (fish) /48 (invertebrates) hours as well as at zero hours.	–	x	x	–	–	–
12c.	Chemical analyses (measurements of test concentration levels) were not performed on the test solutions.	–	–	x	–	–	x
12d.	Unexplained variability between the 0-hour and 96-hour measured concentrations at the two highest test levels (i.e., mean measured concentrations substantially increased)	–	–	–	x	–	–
12e.	Although a static test system was used, the concentrations in the test vessels should have been measured, as other studies with the same test chemical showed that the concentrations varied as much as 30% during a 96-hour exposure period. Therefore, the nominal concentrations did not accurately reflect the true concentrations to which the fish were exposed.	–	–	–	x	–	–
12f.	The test was aerated without chemical analysis of test solutions.	–	–	–	–	x	–
13.	A control group of animals for the inert/carrier ingredients present in the formulation was lacking.	x	x	x	–	–	–
14a.	The acclimation period for the fish prior to initiation of the test was less than half the recommended length of time.	x	x	–	–	–	–

Table A9 continued: Rejection Rate Analysis (US EPA, 1994)

	Rejection reason	Bluegill (Freshwater)	Rainbow (Freshwater)	Invertebrate (Freshwater)	Fish (Estuarine)	Mollusks (Estuarine)	Shrimps (Estuarine)
14b.	Fish acclimation records indicate the test fish were acclimated during the time the definitive study was conducted.	–	x	–	–	–	–
15.	Chemical was recovered in the dilution water control at a level exceeding those in the two lowest test concentrations.	–	x	–	–	–	–
16.	There was very low recovery of the test chemical from the stock solutions at zero and 96 hours. Measurements for the two lowest test concentrations were only given for 96 hours, as the chemical was not detected in the zero-hour samples.	–	x	–	–	–	–
17.	The weights of the test fish exceeded the recommended range.	–	x	–	–	–	–
18a.	The test temperature exceeded that recommended for rainbow trout.	–	x	–	–	–	–
18b.	Temperature of the water was not monitored during the tests.	–	–	x	–	–	–
18c.	No temperature data were provided to indicate if the chambers were monitored at least every 6 hours.	–	–	x	–	–	–
19.	The biological loading of the system was greater than recommended.	–	x	–	–	–	–
20.	Fish mortality during the acclimation period was higher than recommended.	–	x	–	–	–	–
21a.	The dissolved oxygen during the test was supersaturated (over 100% saturation)	–	x	–	–	–	–
21b.	Dissolved oxygen levels fell below 60% saturation at 48 hours and below 40% saturation at 96 hours in two test vessels.	–	x	–	x	–	–
21c.	There were deviations from recommended test solution characteristics. For example, the dissolved oxygen concentration was below recommended values of 60% to 100% saturation.	–	–	–	–	x	–
21d.	Dissolved oxygen levels were below 60% saturation in the last 48 hours of exposure in a flow-through test system.	–	–	–	–	x	–
22.	Organisms were not randomly distributed to test vessels.	–	–	x	–	–	–
23.	Percent a.i. of the tested formulation was not given. The test material was not identified by lot or batch numbers. There was no indication if the concentration used in the study were based on the percent a.i. or total formulated product.	–	–	x	–	–	–
24.	The photoperiod was not 16-hour light/8-hour dark as recommended, but total darkness.	–	–	x	–	–	–
25.	Use of dechlorinated water as a portion of the dilution water is not recommended.	–	–	x	–	–	–
26.	There was insufficient new shell growth in the control oysters to adequately determine the effect of the pesticide on shell deposition.	–	–	–	–	x	–
27.	Raw data on shell deposition were not provided.	–	–	–	–	x	–