Influence of data selection on aquatic ecotoxicity characterization factors for selected persistent and mobile substances

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Influence of data selection on aquatic ecotoxicity characterization factors for selected persistent and mobile substances

Rahul Aggarwal[1] · Hanna Holmquist[2] · Rickard Arvidsson[1] · Efstathios Reppas-Chrysovitsinos[1] · Gregory Peters[1]

Abstract

Purpose This paper addresses the lack of ecotoxicity characterization factors (CFs) for persistent and mobile (PM) chemicals in life cycle assessment. The specific aims are (1) to provide CFs for 64 selected chemicals and benchmark them against the USEtox database, (2) to propose an ecotoxicity data harmonization strategy, and (3) to analyze the influence of ecotoxicity data sources and data harmonization strategies.

Methods Sixty-four per- and polyfluoroalkyl substances (PFAS), triazines, and triazoles were selected. An ecotoxicity data harmonization strategy was developed to ensure data consistency. A screening strategy for including transformation products was also developed. Existing CFs in the USEtox organic substances database (version 2.01) were identified, and new CFs were calculated based on available data sources. The USEtox model (version 2.13) was used for calculations, and the collected ecotoxicity data from different sources were varied to test their influence.

Results and discussion The ecotoxicity data harmonization strategy excluded a considerable amount of incompatible data. To the list of 64 chemicals, 3 transformation products were added. Out of the 67 chemicals, experimental ecotoxicity data were available for 47, leading to the use of in silico tools and average values to fill data gaps for 20 chemicals. CFs for 67 chemicals could thus be provided. Comparing CFs calculated based on experimental and estimated ecotoxicity data reveals that the estimated data leads to considerably different CF values.

Conclusions The paper provides strategies for ecotoxicity data harmonization and the inclusion of transformation products. The limited availability of experimental ecotoxicity data and differing results of the applied estimation methods highlight the need for further development of estimation methods. In addition, further development in including transformation products is recommended, which is particularly relevant for PM chemicals.

Keywords Ecotoxicity · USEtox · Characterization factor · PFAS · Triazines · Triazoles

1 Introduction

Persistent and mobile (PM) chemicals are long-lived in the environment and exhibit considerable mobility in aquatic ecosystems (Neumann and Schliebner 2019). In a recent literature review of PM chemicals by Hale and Kalantzi et al.

✉ Rahul Aggarwal
   rahula@chalmers.se

1   Environmental Systems Analysis, Chalmers University
    of Technology, Vera Sandbergs Allé 8, 41296 Gothenburg,
    Sweden

2   IVL Swedish Environmental Research Institute,
    SE-400 14 Goteborg, Sweden

(2022), 64 specific PM chemicals were identified as potential threats to drinking water due to their chemical properties and widespread use in Europe (Arp & Hale 2022; Hale et al. 2022; Jin et al. 2020). These chemicals belong to the groups per- and polyfluoroalkyl substance (PFAS), triazines, and triazoles, which are used in products such as textile impregnation, fire-fighting foam, ammunition, artificial turf, fungicides, and herbicides (Glüge et al. 2020; Masís-Mora et al. 2019). PFAS are substances containing fluorinated aliphatic carbons (ECHA 2023), such as polytetrafluoroethylene, also known as Teflon. Triazines are nitrogen-containing organic compounds, including the herbicide atrazine. Triazoles are heterocyclic compounds with a five-membered aromatic ring of two carbon atoms and three nitrogen atoms, like the antifungal agent voriconazole (Sahu et al. 2013). When released along the life cycles of

products, these PM chemicals might cause, in addition to drinking water contamination, ecotoxicity impacts, and form toxic degradation products. In this paper, we seek to enable the assessment of the life cycle ecotoxicity impacts of these 64 chemicals.

For this, USEtox is an open-source life cycle impact assessment (LCIA) method for calculating ecotoxicity (and human toxicity, albeit not covered further herein) CFs that hold official endorsement from the UNEP/SETAC Life Cycle Initiative, the European Commission, the World Business Council for Sustainable Development, and the United States Environmental Protection Agency (Fantke et al. 2017). However, calculating CFs with USEtox is data intensive, requiring 18 different parameters for each chemical. Also, USEtox CFs are applied to the emitted compound, potentially neglecting the impact of transformation products (Rosenbaum et al. 2008; Schulze et al. 2001). These degradation products may be highly persistent, mobile, and toxic and should therefore preferably be considered in an LCA along with the parent chemicals (Fenner et al. 2000; Schenker et al. 2007; Van Zelm et al. 2010). Otherwise, the impacts of an emission of the parent chemicals may be underestimated (Boxall et al. 2004). The importance of including ecotoxicity impacts in life cycle assessment (LCA) has been demonstrated (Larsen et al. 2009) and is increasingly done in LCA practice (Rosenbaum 2015).

The ultimate aim of this paper is to provide a practical approach for calculating a medium-large set of ecotoxicity CFs, applied to the selected 64 chemicals. In doing so, the influence of different ecotoxicity data sources and data harmonization strategies on the CFs is explored, including in silico methods to fill data gaps. An ecotoxicity data harmonization strategy and a screening strategy for including the most relevant transformation products are developed. Finally, the calculated CFs are benchmarked against available pre-calculated CFs in the USEtox organic substances database, highlighting differences and similarities stemming from ecotoxicity data selection.

2 Methods

2.1 Overview

The 64 chemicals identified as a potential threat to European drinking water comprised 24 PFAS, 17 triazines, and 23 triazoles are listed in Table S1 in the Supplementary Information (SI). First, the study evaluated the coverage of these chemicals in the USEtox organic substances database (version 2.01). To calculate freshwater ecotoxicity CFs for the chemicals (both those covered and those not covered), the USEtox model (version 2.13) was used (Fantke et al. 2017; Rosenbaum et al. 2008). In USEtox, freshwater

ecotoxicity CFs [$PAF.m^3.d/kg_{emitted}$] are calculated as the product of three factors in a matrix system of fate factors (FF) in $kg.kg^{-1}.d^{-1}$, environmental exposure factors (XF) are unitless, and freshwater ecotoxicological effect factors (EF) in $PAF.m^3.kg^{-1}$:

$$CF = FF \times XF \times EF \tag{1}$$

Data for FF, XF, and EF were collected from a range of publicly accessible sources. Experimental ecotoxicity data were retrieved from CompTox Version 2.1.1, having toxicity data from ToxValDB v9.1.1 (*CompTox* 2022; Williams et al. 2017). In total, 5002 ecotoxicity data points were retrieved, covering 15 PFAS, 12 triazines, and 21 triazoles. These data points were subjected to the developed harmonization strategy described in Sect. 2.3. Estimated ecotoxicity data were retrieved from different quantitative structure–activity relationships (QSARs), including ECOSAR™ Version 1.11 [$n = 39$] accessed through EPI Suite v4.11 (U.S.EPA 2023b), US EPA TEST [$n = 44$] (U.S.EPA 2020), VEGA [$n = 5$] (*VEGA* HUB 2022), and Danish (Q)SAR database [$n = 50$] (*Danish (Q)SAR* 2022) with aquatic toxicity from DTU-developed models (Battery, Leadscope and SciQSAR). Physio-chemical data was retrieved from different sources, such as molecular weight (MW) through Chemspider (Chemspider 2022); pKa chemical class (neutral, acid, base, and amphoter), pKa.gain, and pKa.loss through Chemaxon (Chemaxon 2022); and K_{OW}, K_{OC}, KH25C, Pvap25, Sol25, kdegA, kdegW, kdegSd, and kdegSl through EPI Suite v4.11(U.S.EPA 2023b).

In a recent publication, Owsianiak et al. (2023) provide new recommendations for ecotoxicity characterization, stemming from collaborative efforts by the Ecotoxicity Task Force and the SETAC Pellston Workshop. These recommendations include applying EC10 equivalents to calculate EFs instead of EC50 values as in the original USEtox framework. However, as these recommendations are not adopted by USEtox yet, this study uses USEtox (version 2.13). In addition, a PFAS-adapted USEtox model was previously developed by Holmquist et al. (2020) based on USEtox (version 2.1) to address the limitations of USEtox for calculating CFs for PFAS (Owsianiak et al. 2023). While this PFAS-adapted model can be expected to calculate more relevant CFs for PFAS in general, it also requires more input data (31 items as compared to 18 in the original model), which are difficult to obtain. To quantify the influence of this model adaptation, both the PFAS-adapted model and USEtox (version 2.13) were applied to calculate CFs for three PFAS compounds for which data were available in the PFAS-adapted model: perfluorooctanoic acid (PFOA), perfluorohexanoic acid (PFHxA), and perfluorobutanesulfonic acid (PFBS) discharged to the freshwater compartment (disregarding the species richness and

groundwater recirculation). The results revealed an average change in CFs of only 2.5% (ranging from 0 to 7%), showing that the PFAS-adapted model and USEtox (version 2.13) provide similar CFs for these three PFAS. The potential to generalize conclusions is limited since only three substances were included in the comparison. However, considering these modest differences and the substantial additional data requirement of the PFAS-adopted model, we conclude that using USEtox version 2.13 without PFAS adaptations is justified in this work.

2.2 Transformation product screening strategy

Including every possible transformation byproduct from the 64 chemicals is infeasible. Therefore, a simplified in silico screening strategy was adopted to select the most relevant transformation products in terms of persistence and ecotoxicity. First, potential transformation products were predicted for each parent chemical using two prediction tools: the United States Environmental Protection Agency's Chemical Transformation Simulator (CTS) (Wolfe et al. 2016; Yuan et al. 2021) and enviPath (enviPath 2022; Wicker et al. 2016). Then, a persistence check was conducted by comparing the half-life of the parent chemical and three generations of its transformation products in water to exclude all transformation products that were less persistent than their respective primary chemical. The EPI Suite level III fugacity model with a fixed temperature of 25 °C was used to estimate the half-life in water (Aronson et al. 2006; U.S.EPA 2023b). The remaining transformation products were subjected to an ecotoxicity check by comparing predicted concentration that is lethal to 50% of the test species (LC50) for the fathead minnow for an exposure duration of 96 h of the parent and its respective transformation products, to exclude all the transformation products that were less toxic than their respective parent. To estimate ecotoxicity, the EPA Toxicity Estimation Software Tool (TEST) was applied (U.S.EPA 2020). Transformation products with greater or equal persistence and toxicity than their respective parent were added to the list of the 64 chemicals. This procedure is described in more detail in Sheet S7 in the SI.

An alternative approach to including relevant transformation products as separate chemicals would be to increase the parent compound CFs in proportion to the CFs of its transformation products as proposed by Van Zelm et al. (2010), acknowledging that the net impact of a chemical will be the sum of its impact and the impact of its transformation products. However, this is not done in this study since the estimation of transformation products here is not quantitatively stoichiometric.

2.3 Data harmonization strategy

For ecotoxicological effect data, a data harmonization strategy was developed since CFs have been shown to be particularly sensitive to this parameter (Holmquist et al. 2020; Roos et al. 2017). The steps of this strategy are provided below in brief and explained in more detail in Sheet S6 in the SI:

1. **Chemical identifiers**: Chemical Abstracts Service (CAS) registration numbers were matched with the respective reported chemical
2. **Ecosystem type**: Freshwater data points were included, excluding data points referring to other ecosystems
3. **Data point qualifiers**: Data points with numeric qualifiers such as $>$, $\geq$, $<$, and $\leq$ were excluded
4. **Data point unit**: Units of all data points were converted into mg/L, excluding data points unconvertible to this unit
5. **Endpoint classification**: Data points endpoints were classified into NOEC eq, EC10 eq, EC50 eq, and LOEC eq, excluding data points referring to other endpoints
6. **Test species name**: Tested species were assigned common species names as per US EPA ECOTOX (U.S.EPA 2023a)
7. **Species grouping**: Data points of the tested species were classified into seven species groups: (i) algae and cyanobacteria, (ii) aquatic plants (other than algae), (iii) bacteria, (iv) fish, (v) crustacean invertebrates, (vi) other invertebrates, and (vii) vertebrates (Aurisano et al. 2019)
8. **Exposure duration**: Exposure duration units of the data points were converted into days
9. **Acute/chronic exposure duration threshold**: The chronic exposure duration threshold considered was 1 day for algae, cyanobacteria, bacteria, and fungi; 4 days for crustacean invertebrates; and 7 days for fish, other invertebrates, vertebrates, and aquatic plants other than algae (Aurisano et al. 2019; Payet 2004)
10. **Acute/chronic based on exposure duration**: Data points were categorized as either acute or chronic by comparing their exposure duration with their respective chronic exposure duration threshold
11. **Species life stages and sensitivity**: Test species' life stages were categorized as either sensitive or non-sensitive. It was also noted whether the test was reportedly chronic (e.g., chronic reproduction and chronic development)
12. **Acute/chronic based on life stages combined with risk assessment class**: Data points categorized as acute based on their respective exposure duration threshold were considered. All acute data points from reportedly

chronic tests with sensitive life stages were categorized as chronic, as a precautionary approach

13. **Source availability**: Data points with unavailable original sources were excluded

14. **Data type**: Each data point was classified as either experimental or estimated. Estimated data points consisting of QSAR generated results, expert judgment, and read-across data were excluded from the experimental dataset

15. **Study reliability**: The Klimisch score of the included data points was either 1 (reliable without restriction) or 2 (reliable with restrictions). Higher scores were excluded (Klimisch et al. 1997). In the absence of a Klimisch score, expert judgment was used to evaluate the reliability

16. **Endpoint harmonization**: Endpoints were converted into chronic EC50

17. **Ecologically relevant effects**: The relevance of the effects for aquatic ecosystems was checked, excluding data points with unclear relevance

18. **Most sensitive effect**: Data points with the most sensitive effect per study, per chemical, and per test species were identified and included, excluding other data points

19. **Duplicate removal**: Duplicate data points per chemical were removed

2.4 Benchmarking of characterization factors

To understand the relative freshwater ecotoxicity potential of the chemicals in this study compared to all the other chemicals in the USEtox 2.01 organic substances database, the harmonized datasets for PFAS, triazines, and triazoles are ranked against the USEtox 2.01 dataset. The study also examines correlations between being in a particular group of PM chemicals, such as PFAS, triazines, and triazoles, and related toxicity potential. Additionally, the calculated EFs of the 18 chemicals already covered in the database were compared to their pre-calculated USEtox EFs in the USEtox organic substances database. The robustness of the relationship between our calculated EFs and pre-calculated USEtox EFs was determined through the coefficient of determination (R^2) of log-transformed fit regression analysis.

2.5 Alternative data harmonization strategies

In our baseline strategy, all steps proposed in the data harmonizing strategy were followed. However, in strategy 1, all data points with unacceptable numeric qualifiers, such as $>, \geq, <,$ and $\leq$, were included, as done by other authors (Aurisano et al. 2023). All such qualifiers were then interpreted as "$=$". In strategy 2, all data points were assumed to be acute and extrapolated to a chronic equivalent (a

conservative approach). In strategy 3, step 18 of the harmonizing scheme was removed, assuming all effects and endpoints as equally sensitive and relevant, respectively. Then, the correlation between the calculated EFs from the alternative strategies and the EFs from the baseline harmonizing strategy was assessed using the R^2 of log-transformed fit regression analysis. These alternative strategies were implemented to investigate the influence of disregarding certain data- and time-intensive steps on the results. The purpose of strategies 1 and 3 was to examine the influence of a more data-inclusive curation process. Strategy 2 removed the need to classify data as acute or chronic.

3 Results and discussion

This section first describes the identification of transformation products added to the list of chemicals. Second, it outlines the outcome of the ecotoxicity data collection and harmonization. Third, it details the coverage and calculation of CFs along with a comparative benchmarking against the USEtox database. Fourth, the section presents the outcome of applying alternative data harmonization strategies, followed by a final comparison of estimated versus experimental data.

3.1 Identified transformation products

Three hundred twenty-two transformation products were identified; 166 from CTS and 156 from enviPath. In the persistency check, EPI Suite calculated half-lives for 384 out of 386 chemicals, resulting in 78 transformation products being equally persistent and 242 being less persistent than their parent compounds. Only one compound (2-[(4-chlorophenyl)methyl]-5-(2-hydroxypropan-2-yl)-1-[(1H-1,2,4-triazol-1-yl)methyl]cyclopentan-1-ol) was more persistent than its parent (ipconazole). In the toxicity check, TEST estimated the ecotoxicity of 162 out of 386 chemicals, resulting in eight transformation products being more toxic and 109 being less toxic than their respective parent. However, for 19 parents (one PFAS, nine triazines, and nine triazoles), TEST could not estimate a toxicity value. Out of the 79 transformation products identified as having equal or greater persistence, only three were more toxic: difenoconazole-ketone, 1-(2,4-dichlorophenyl)-2-(1H-1,2,4-triazol-1-yl)ethenone, and 1-[(2Z)-3-(2-Chlorophenyl)-2-(4-fluorophenyl)prop-2-en-1-yl]-1H-1,2,4-triazole. These three were thus added to the list for data collection. However, as no experimental data were available for these chemicals, only estimated values from ECOSAR™ Version 1.11 [$n=3$] accessed through EPI Suite v4.11 (U.S.EPA 2023b) and US EPA TEST [$n=3$] (U.S.EPA 2020) were included in calculating their CFs. As this study departs from a list of

water supply-relevant contaminants, which already includes several transformation products alongside parent chemicals, the addition of chemicals to this list was relatively limited and this procedure is expected to have additional relevance in assessments of PM chemicals in commercial products.

3.2 Data collection and harmonization results

The data harmonization strategy in Sect. 2.3 was applied to the ecotoxicity data retrieved from CompTox Version 2.1.1. In total, 5002 experimental ecotoxicity data points were retrieved, covering 600 data points for 15 PFAS, 2586 data points for 12 triazines, and 1816 data points for 21 triazoles. The number of data points available per chemical ranges from only 1 for 1H,1H,2H,2H-perfluorohexanesulphonic acid to 1540 for atrazine. Data points availability per chemical for different chemical groups also vary, with an average and standard deviation of 40 and 61 for PFAS, 216 and 365 for triazines, and 86 and 88 for triazoles.

No data point was excluded at step 1, but step 2 resulted in exclusion of 2160 data points related to the terrestrial ecosystem. Step 3 resulted in the exclusion of 250 data points with uncertain numeric qualifiers: $>$ [$n = 208$], $\geq$ [$n = 3$], $<$ [$n = 37$], and $\leq$ [$n = 2$]. The remaining data points [$n = 2495$] were checked for consistent units of mg/L in step 4, resulting in the exclusion of 96 data points with unconvertible units (g/ha, mg/kg-day, mg/kg bdwt, mg/kg food, etc.). In step 5, data points were classified under four predefined effect endpoints (NOEC eq; EC10 eq; EC50 eq; and LOEC eq), resulting in the exclusion of 203 data points with unclassified endpoints (EC25, EC90, IC25, LC90, NR-LETH, etc.). From step 6 to 14, only one data point was excluded at step 8. Step 15 resulted in the exclusion of 22 data points with Klimisch score of 3 or 4. Steps 16 and 17 resulted in no exclusions but in step 18, 1080 data points with either less sensitive effects or less relevant endpoints (in individual comparisons within the dataset of single studies) were excluded. At step 19, there were no duplicates remaining, resulting in no exclusion.

After harmonization, the final set contained 1189 data points; 174 points for 14 PFAS, 668 points for 12 triazines, and 347 points for 21 triazoles. There was a loss of one PFAS chemical, 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)-propanoic acid, having initially seven data points, but five were excluded at step 2 (belonging to the terrestrial ecosystem) and two were excluded at step 3 with uncertain numeric qualifiers ">". Harmonized data point availability per chemical ranged from one for perfluorohexanesulfonic acid, perfluoroheptanoic acid, 2,2,3-Trifluoro-3-(1,1,2,2,3,3-hexafluoro-3-(trifluoromethoxy)propoxy) propanoic acid, and 1H,1H,2H,2H-Perfluorohexanesulphonic acid, to 487 for atrazine. Data point availability per chemical also varied, with an average and standard deviation of 12 and 18 for PFAS, 54 and 116 for triazines, and 16 and 21 for triazoles.

In terms of the number of species and species groups, the final experimental harmonized dataset includes an average of 10 species and three species groups per chemical, with an average of five species and two species groups for the 14 PFAS compounds, 20 species and four species groups for the 12 triazines, and nine species and four species groups for the 21 triazoles. The number of species per chemical ranges from one for five PFAS compounds and one triazine to 140 for atrazine. Additionally, the number of species groups also varies, ranging from one for six PFAS compounds and one triazine to seven for tebuconazole.

The data encompasses a wide temporal range, spanning from 1965 to 2020. The harmonized dataset includes data points distributed as follows: 58% after 2004 and 35% after 2010. The year 2004 is significant because it was the cut-off year for most of the data included in the current USEtox database. The distribution varies for different chemical groups. For PFAS, the distribution is 90% after 2004 and 77% after 2010, indicating that a significant portion of the data for PFAS was produced in the last decade. For triazines, the distribution is 44% after 2004 and 22% after 2010, so most of that data is older than two decades. For triazoles, the distribution shows 69% after 2004 and 39% after 2010, so most of that data was produced within the last two decades.

To investigate the variability of QSAR estimates in CF calculations and to fill experimental data gaps, the combined estimated ecotoxicity data retrieved from ECOSAR™ Version 1.11 [$n = 39$], US EPA TEST [$n = 44$], and Danish (Q) SAR database [$n = 50$] (Danish (Q)SAR Database 2022) were harmonized according to the proposed harmonizing scheme in Sect. 2.3. VEGA data within applicability range are only available for 5 chemicals so not used in this study. ECOSAR™ Version 1.11 covers the species fish, daphnid, and green algae, US EPA TEST covers daphnia and fathead minnow, and the Danish (Q)SAR database covers daphnia magna, fathead minnow, and *Pseudokirchneriella subcapitata*. The final estimated harmonized data contained 524 data points covering 161 data points for 21 PFAS, 158 data points for 17 triazines, and 205 data points for 21 triazoles. In terms of the number of species and species groups, the overall average was four species and three species groups per chemical, with an average of three species and two species groups for the 21 PFAS, five species and four species groups for the 17 triazines, and four species and four species groups for the 21 triazoles. All data can be found in the SI, Table S5.

3.3 Characterization factor calculation and benchmarking

This study evaluated the coverage of the 67 chemicals in the USEtox organic substances database (version 2.01), showing that USEtox currently covers none of the 24 PFAS, 10/17 triazines, and 8/23 triazoles, resulting in an overall coverage

of 30% (18/64). This finding indicates a lack of CFs for PM substances relevant for drinking water systems especially PFAS. The 67 CFs calculated are shown in Fig. 1 and listed in Table S2 in the SI. As can be seen, they range almost six orders of magnitude. Nearly 70% of the CFs (47/67) were calculated using experimental data (14 PFAS, 12 triazines, and 21 triazoles), implying an experimental data gap of 30%. To address this data gap, estimated ecotoxicity data from QSARs were obtained for an additional 17 chemicals (8 PFAS, five triazines, one triazole, and three transformation products). Then, a data gap remained for three chemicals (two PFAS and one triazole). To fill those data gaps for the two PFAS, a simplified regression analysis was explored as explained in Table S4 in the SI, relating the number of perfluorinated carbons to the calculated CFs as shown in Figure S1. However, as the correlation was low, these values were not used in this study. Instead, the average of the other 22 PFAS was used to fill the data gaps for the two last PFAS. For the single remaining triazole, a CF was also calculated using the average CF of the other 22 triazoles.

The CFs of the transformation products of difenoconazole-ketone, 1-(2,4-dichlorophenyl)-2-(1H-1,2,4-triazol-1-yl) ethenone, and 1-[(2Z)-3-(2-chlorophenyl)-2-(4-fluorophenyl)prop-2-en-1-yl]-1H-1,2,4-triazole are 9%, 7%, and 13% of their respective parent chemicals. One might have expected the transformation product CFs would be larger given the initial assessment that identified them as more persistent and toxic, but the more comprehensive calculation after full data curation for these three chemicals means that

additional ecotoxicity data for three species were included in the calculation, rather than the single species (fathead minnow) considered in the initial assessment. Nevertheless, in absolute terms, the CFs for these three range from 3×10^3 to 4×10^4 PAF.m^3.d.kg^{-1}, which places them in the center of the range in Fig. 1.

The calculated CFs were benchmarked against the precalculated CFs in the USEtox organic substances database (version 2.01) [$n = 2499$]. To illustrate this variability, box plots depicting the range of CFs for different PM chemical groups and the USEtox organic substances database are presented in Figure S4 in Sheet S8 in the SI. These plots provide a visual representation of the distribution of CFs, highlighting the wide range of values for each chemical group and that all chemicals considered here are within the range of CFs in the USEtox database.

In this study, EFs were calculated with experimental data for 18 chemicals that were already present in the USEtox database. These EFs were compared to the original EFs from the USEtox organic substances database (version 2.01), as shown in Fig. 2. The regression analysis yielded an R^2 of 0.63, indicating a moderate correlation. The corresponding regression analysis for CFs yielded an R^2 of 0.69, also indicating a moderate correlation. Nevertheless, when exclusively altering the ecotoxicity values of chemicals shared between this study and the original USEtox dataset (keeping all other fate and exposure-related data constant), the R^2 notably improved to 0.99. This observation implies that the fluctuations in CFs are predominantly related to differences
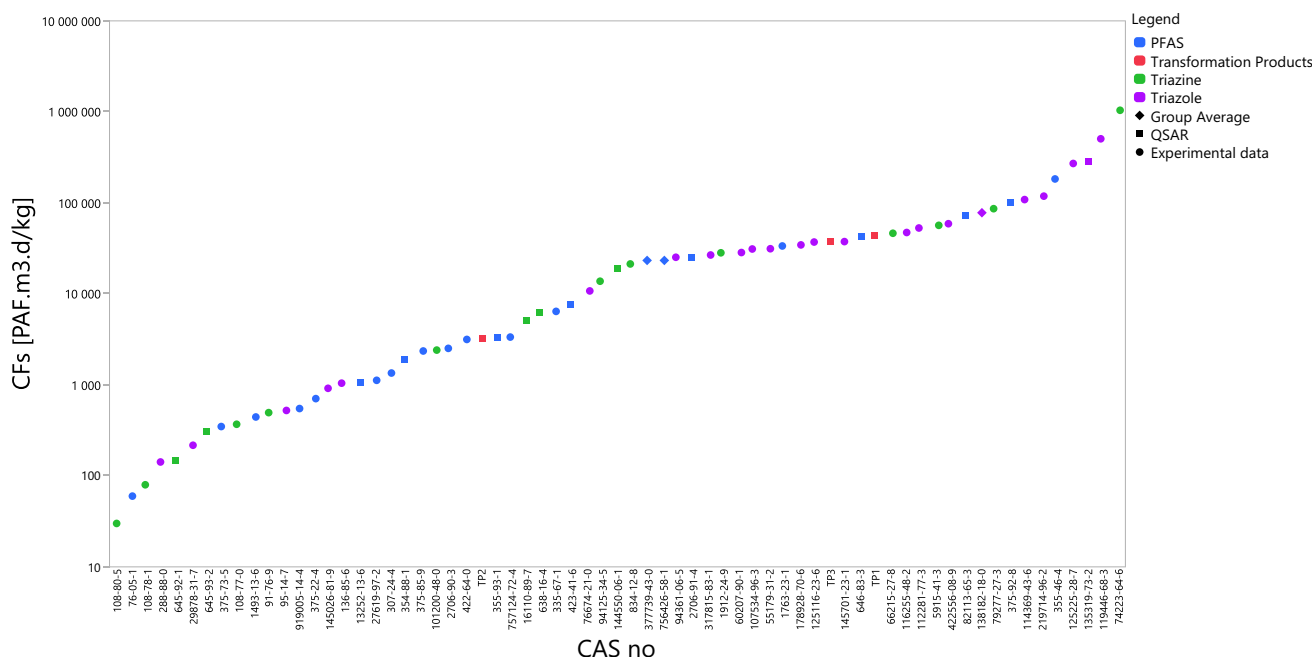

Fig. 1 Calculated freshwater ecotoxicity CFs for 67 chemicals (24 PFAS, 17 triazines, 23 triazoles, and 3 transformation products). Forty-seven CFs were calculated using experimental data, 17 with QSAR data, and 3 with group averages

Fig. 2 Log-transformed fit regression analysis of pre-calculated USEtox version 2.13 database EFs versus calculated EFs with experimental ecotoxicity data from this study (n = 18, R^2 = 0.63)
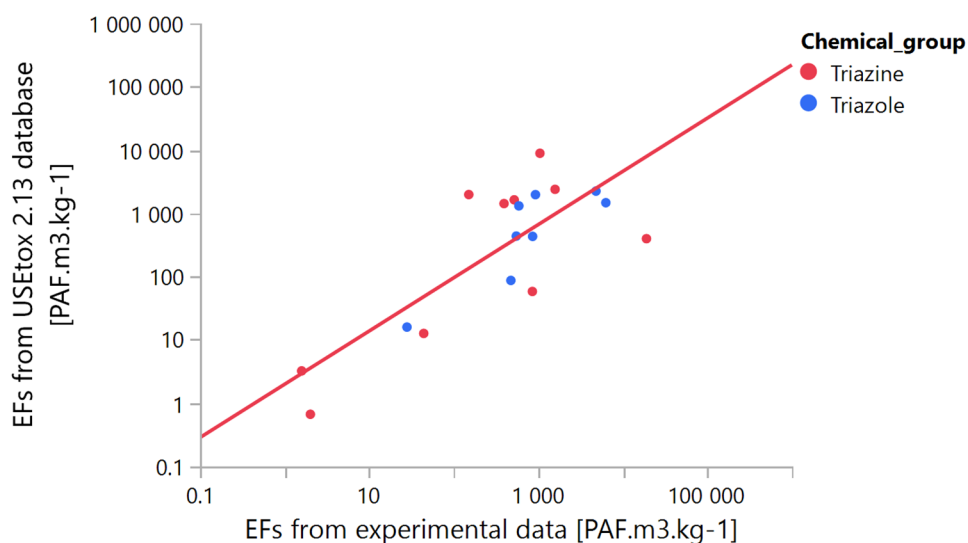
in ecotoxicity values rather than other input data, aligning with previous findings (Holmquist 2020; Roos et al. 2017).

The original USEtox ecotoxicity factors were calculated based on two databases with only EC50 values, namely, acute EC50 values from the e-toxBase database of the National Institute for Public Health and the Environment (RIVM), accessed through van Zelm et al. (2009) and Zelm et al. (2007). Furthermore, chronic and acute EC50 data from ECOTOX and IUCLID accessed through Payet (2004) were included, applying an acute-to-chronic ratio to derive the chronic values (Fantke et al. 2015). Therefore, benchmarking the CFs calculated in this study against the pre-calculated USEtox CFs serves as a temporal sensitivity test, considering the addition of new ecotoxicity data. For 18 chemicals, 51% data points were obtained since 2005 and 28% since 2011. This highlights that including new ecotoxicity data can notably influence CFs due to the dominant influence of EFs in the CF calculations. The original USEtox factors were published with ecotoxicity data sets that are now more than a decade old, and the new data generated since then, incorporated in the CFs calculated herein, has contributed to changing the CFs. Furthermore, since an EF is based on an average ecotoxicity among species, adding data for more species should make the EF more representative of a range of species.

3.4 Influence of data sources and harmonization

In the baseline data harmonization strategy, the final harmonized data consisted of 1189 data points. After harmonization, strategy 1 included 1336 points for 48 substances (198 points for 15 PFAS, 747 points for 12 triazines, and 391 points for 21 triazoles). Strategy 2 included 1215 points for 47 substances (180 points for 14 PFAS, 687 points for 12 triazines, and 348 points for 21 triazoles). The increase in data points compared to the baseline strategy stems from step 18, where the preferred order was modified to EC50 eq acute > EC10 eq acute > LOEC eq acute > NOEC eq acute in the absence of chronic data. As availability of acute data tends to be higher than chronic data, this modification resulted in more data inclusions compared to the baseline strategy. Strategy 3 included 2214 data points for 47 substances (342 data points for 14 PFAS, 1246 data points for 12 triazines, and 626 data points for 21 triazoles).

The calculated EFs [n = 47] from different strategies were correlated with EFs calculated using the baseline harmonizing strategy. In strategy 1, the R^2 value was 0.94, indicating a strong correlation as shown in Fig. 3. This suggests that the removal of unacceptable numeric qualifiers had low influence on most of the EFs in this study that may not be valid always. In strategy 2, the R^2 was 0.99, indicating a strong correlation. This suggests that the classification of acute and chronic effects did not notably affect most EFs (note that this could be different if the data harmonization scheme would only include chronic data). In strategy 3, the R^2 was 0.94, again indicating a strong correlation. This suggests that considering only sensitive effects and relevant endpoints does not have a notable influence on most of the EFs.

The final harmonized dataset consisted of 42% chronic and 58% acute data points. The acute data were extrapolated to chronic equivalents as per step 16 before calculating CFs. The distribution varied across different chemical groups, with PFAS having 60% acute and 40% chronic data, triazines having 55% acute and 45% chronic data, and triazoles having 62% acute and 38% chronic data. Thus, the majority of experimental data was based on acute tests. This was not surprising as chemical hazard testing almost always starts with acute tests, which are then followed up with chronic
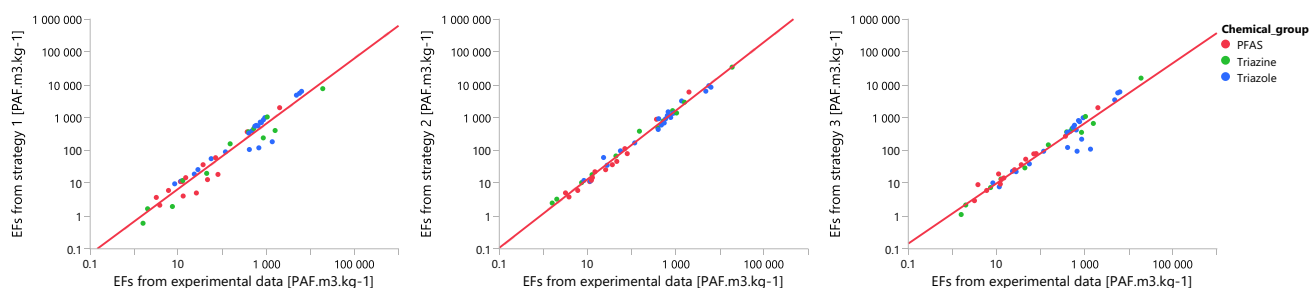
Fig. 3 Log-transformed fit regression analysis of calculated EFs for three harmonization strategies (strategy 1: including data points with unacceptable numeric qualifiers, strategy 2: assuming all data points to be acute and extrapolated to a chronic equivalent, and strategy 3: assuming all effects and endpoints as sensitive and relevant) against the baseline strategy using all current data revealing a strong correlation in strategy 1 (n=47, R^2=0.94), strategy 2 (n=47, R^2=0.99), and strategy 3 (n=47, R^2=0.94)

effect testing. However, extrapolation from acute to chronic effects adds uncertainty to the CF.

3.5 Comparing QSAR models with experimental data

To compare EFs based on experimental versus combined estimated ecotoxicity data, a regression was made between the EFs calculated with experimental data and EFs based on ecotoxicity data estimated by combining the results of the five QSARs [$n=45$]. The resulting R^2 was 0.37, as shown in Fig. 4, indicating a weak correlation. The calculation suggested that the QSARs-derived EFs are within two orders of magnitude of the experimental values for most of the chemicals. Furthermore, the R^2 varies notably for different chemical groups. For PFAS [$n=13$], the R^2 was 0.0006, indicating no correlation. In contrast, for triazines [$n=12$], the R^2 was 0.31, indicating only weak correlation. Lastly, for triazoles [$n=20$], the R^2 was 0.50, indicating moderate correlation. These findings reflect the relatively low ability

of the QSARs to provide accurate ecotoxicity data to the CF calculations in this study.

To assess the relative reliability of EFs derived from different QSAR models, the EFs calculated with the two QSARs that correlated best with the experimental EFs were considered individually. (Results for all five QSARs are provided in Table S1 in SI). For the ECOSAR QSAR of EPI Suite [$n=33$], the resulting R^2 value was 0.36, as shown in Fig. 5. This indicates a weak correlation between the calculated EFs and the experimental EFs from ECOSAR. For different chemical groups, the R^2 varied considerably. The R^2 cannot be calculated for PFAS due to a low number of data points [$n=2$]. For triazines [$n=12$], the R^2 was 0.14, indicating a very weak correlation. For triazoles [$n=19$], the R^2 was 0.46, suggesting a moderate correlation.

For the TEST QSAR [$n=33$], the R^2 was 0.53, indicating a better correlation between the EFs calculated with estimated data and those based on experimental data compared to ECOSAR. The R^2 was 0.29 for PFAS [$n=12$] and 0.66

Fig. 4 Log-transformed fit regression analysis of calculated EFs based on estimated versus experimental ecotoxicity data (n=45, R^2=0.37). For different chemical groups, there is no correlation for PFAS (n=13, R^2=0.0006), weak correlation for triazines (n=12, R^2=0.31), and a moderate correlation for triazoles (n=20, R^2=0.50). Dashed lines show 95% confidence intervals
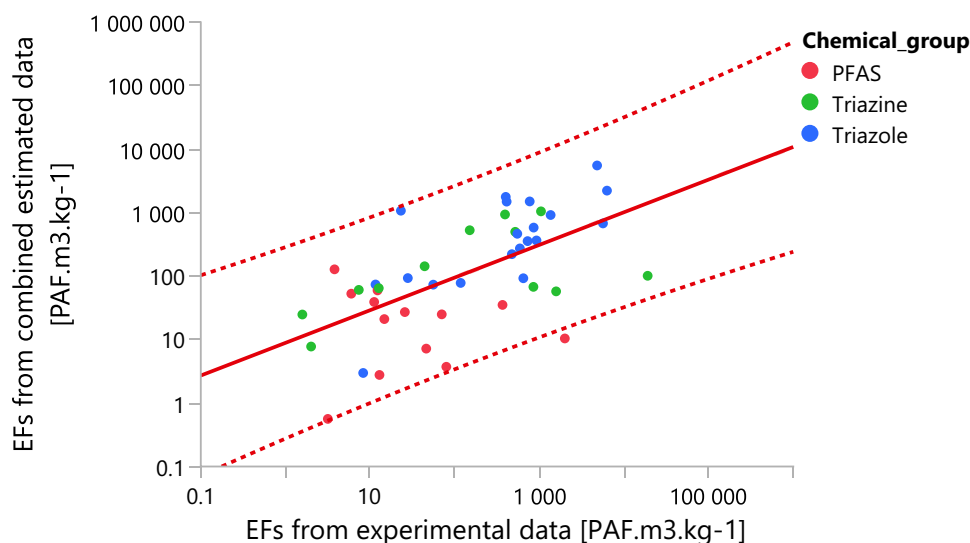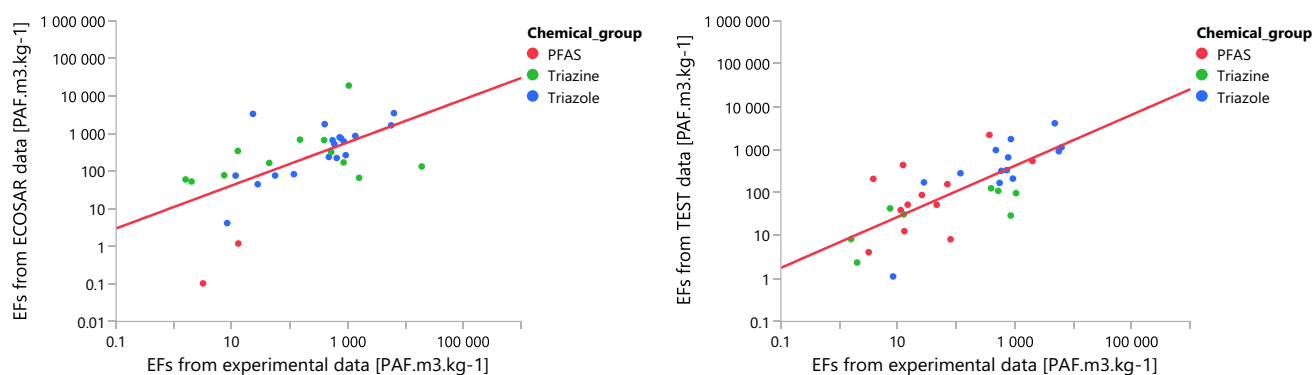
Fig. 5 Log-transformed fit regression analysis of calculated EFs with ecotoxicity data from different in silico tools (ECOSAR and TEST) versus the experimental ecotoxicity data. ECOSAR: $n = 33$, $R^2 = 0.36$. TEST: $n = 33$, $R^2 = 0.53$

for triazoles [$n = 13$]. These regression results suggest that TEST offers greater reliability for calculating EFs for the selected chemicals than the other QSARs examined.

A "species selection bias" might arise due to the limited number of species covered by QSAR models (algae, daphnid, and fish for ECOSAR (not further specified); fathead minnow and *Daphnia magna* for US EPA TEST, *pseudokirchneriella* green algae for the Danish (Q)SAR). Differences in coverage also include both type of endpoints and test duration. Considering these differences in the coverage of species and effects, the low correlation obtained between experimental and estimated EFs might not reflect the general performance of the QSARs.

4 Conclusions

In this paper, we followed an approach through which ecotoxicity CFs for 67 chemicals could be provided, including 49 previously not characterized. The benchmarking with pre-calculated CFs from the USEtox database demonstrates that these CFs fall within the range of CFs currently present in the USEtox database. Additionally, comparing CFs for chemicals present in the original USEtox database with new calculations in this study demonstrates the influence of including up-to-date toxicity data. The new data influenced the CFs notably and underscore the importance of continuous toxicological research and updating of toxicity data in databases to ensure more reliable ecotoxicity CFs. When comparing EFs based on experimental and estimated ecotoxicity data, a weak correlation emerged. This implies a modest ability of the QSARs to provide accurate ecotoxicity data to the CF calculations in this study. Further development of QSAR methods that can provide ecotoxicological data relevant for CF calculations is therefore required.

This paper's focus on ecotoxicity led to the development of an ecotoxicity data harmonization strategy to ensure consistency and transparency during CF calculations. Ecotoxicity data are a critical input parameter, but even after obtaining such data, harmonizing it to the USEtox format can be difficult. Such data are collected for different purposes, such as chemical risk assessment, leading to variations in endpoints, units, and effects. Unless harmonized, inconsistencies in underlying ecotoxicity data can propagate into the calculated CFs, or data gaps can arise if inconsistent data are omitted instead of harmonized. Alternative data harmonization strategies were considered in this study, involving the inclusion and exclusion of numeric qualifiers, classification of acute or chronic exposure durations, and the classification of effects and endpoints as sensitive and relevant. The results obtained from these different harmonization strategies show a high correlation, indicating that the details of the harmonization strategy do not alter the relative ranking of CFs notably.

We also developed a screening strategy to include the most important transformation products from the original 64 chemicals. Since several of the parent chemicals considered are already transformation products, few additional transformation products with higher persistence and toxicity were identified in the screening. However, if the screening strategy would be applied to PM substances released from commercial products, we expect a larger number of identified transformation products. Nevertheless, LCIA requires further development regarding the inclusion of transformation products along with their parent chemicals. This is particularly relevant for PM chemicals considering how these compounds can persist in the environment long after the total degradation of their parent compounds.

Declarations

Conflict of interest The authors declare no conflict of interests.

References

Aronson D, Boethling R, Howard P, Stiteler W (2006) Estimating biodegradation half-lives for use in chemical screening [Article]. Chemosphere 63(11):1953–1960. https://doi.org/10.1016/j.chemosphere.2005.09.044

Arp HPH, Hale SE (2022) Assessing the persistence and mobility of organic substances to protect freshwater resources. ACS Environmental Au. https://doi.org/10.1021/acsenvironau.2c00024

Aurisano N, Albizzati PF, Hauschild M, Fantke P (2019) Extrapolation factors for characterizing freshwater ecotoxicity effects. Environ Toxicol Chem 38(11):2568–2582. https://doi.org/10.1002/etc.4564

Aurisano N, Jolliet O, Chiu WA, Judson R, Jang S, Unnikrishnan A, Kosnik MB, Fantke P (2023) Probabilistic points of departure and reference doses for characterizing human noncancer and developmental/reproductive effects for 10,145 chemicals. Environ Health Perspect 131(3):037016. https://doi.org/10.1289/EHP11524

Boxall AB, Sinclair CJ, Fenner K, Kolpin D, Maund SJ (2004) Peer reviewed: when synthetic chemicals degrade in the environment. Environ Sci Technol 38:368A

Chemaxon (2022) ChemAxon JChem. Retrieved February from https://chemaxon.com/calculators-and-predictors

Chemspider (2022) Retrieved March from http://www.chemspider.com/

CompTox Chemicals Dashboard Version 2.1.1. (2022) Retrieved September from https://comptox.epa.gov/dashboard/batch-search

Danish (Q)SAR Database (2022) Research group for chemical risk assessment and GMO, National Food Institute, Technical University of Denmark. Retrieved March from http://qsar.food.dtu.dk

ECHA (2023) European Chemicals Agency, ECHA. Retrieved March from https://echa.europa.eu/about-us

enviPath (2022) enviPath. Retrieved October from https://envipath.org/

Fantke P, Huijbregts MAJ, Margni M, Hauschild MZ, Jolliet O, McKone TE, Rosenbaum RK, Van De Meent D (2015) USEtox 2.0 user manual (Version 2). https://usetox.org/

Fantke P, Bijster M, Guignard C, Hauschild M, Huijbregts M, Jolliet O, Kounina A, Magaud V, Margni M, McKone TE, Posthuma L, Rosenbaum R, Van de meent D, Zelm R (2017) USEtox 2.0 : documentation (Version 1). https://usetox.org/

Fenner K, Scheringer M, Hungerbühler K (2000) Persistence of parent compounds and transformation products in a level IV multimedia model. Environ Sci Technol 34(17):3809–3817. https://doi.org/10.1021/es0000347

Glüge J, Scheringer M, Cousins IT, DeWitt JC, Goldenman G, Herzke D, Lohmann R, Ng CA, Trier X, Wang Z (2020) An overview of the uses of per- and polyfluoroalkyl substances (PFAS) [https://doi.org/10.1039/D0EM00291G]. Environ Sci Process Impacts 22(12), 2345–2373. https://doi.org/10.1039/D0EM00291G

Hale SE, Kalantzi OI, Arp HPH (2022) Introducing the EU project ZeroPM: zero pollution of persistent, mobile substances [Editorial]. Environ Sci Eur 34(1), Article 108. https://doi.org/10.1186/s12302-022-00681-5

Hale SE, Neumann M, Schliebner I, Schulze J, Averbeck FS, Castell-Exner C, Collard M, Drmač D, Hartmann J, Hofman-Caris R, Hollender J, de Jonge M, Kullick T, Lennquist A, Letzel T, Nödler K, Pawlowski S, Reineke N, Rorije E, . . . Arp HPH (2022) Getting in control of persistent, mobile and toxic (PMT) and very persistent and very mobile (vPvM) substances to protect water resources: strategies from diverse perspectives [Article]. Environ Sci Eur 34(1), Article 22. https://doi.org/10.1186/s12302-022-00604-4

Holmquist H, Fantke P, Cousins IT, Owsianiak M, Liagkouridis I, Peters GM (2020) An (eco)toxicity life cycle impact assessment framework for per-and polyfluoroalkyl substances [Article]. Environ Sci Technol 54(10):6224–6234. https://doi.org/10.1021/acs.est.9b07774

Holmquist, H (2020) Chemical substitution with a life cycle perspective: the case of per- and polyfluoroalkyl substances in durable water repellents. isbn:978-91-7905-299-7

Jin B, Huang C, Yu Y, Zhang G, Arp HPH (2020) The need to adopt an international PMT strategy to protect drinking water resources. Environ Sci Technol 54(19):11651–11653. https://doi.org/10.1021/acs.est.0c04281

Klimisch HJ, Andreae M, Tillmann U (1997) A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25(1):1–5. https://doi.org/10.1006/rtph.1996.1076

Larsen HF, Hansen MS, Hauschild M (2009) Life cycle assessment of offset printed matter with EDIP97: how important are emissions of chemicals? [Article]. J Clean Prod 17(2):115–128. https://doi.org/10.1016/j.jclepro.2008.03.006

Masís-Mora M, Lizano-Fallas V, Tortella G, Beita-Sandí W, Rodríguez-Rodríguez CE (2019) Removal of triazines, triazoles and organophosphates in biomixtures and application of a biopurification system for the treatment of laboratory wastewaters. Chemosphere 233:733–743. https://doi.org/10.1016/j.chemosphere.2019.06.001

Neumann M, Schliebner I (2019) Protecting the sources of our drinking water: the criteria for identifying persistent, mobile and toxic (PMT) substances and very persistent and very mobile (vPvM) substances under EU Regulation REACH (EC) No 1907/2006 | Umweltbundesamt. https://www.umweltbundesamt.de/en/publikationen/protecting-the-sources-of-our-drinking-water-the

Owsianiak M, Hauschild MZ, Posthuma L, Saouter E, Vijver MG, Backhaus T, Douziech M, Schlekat T, Fantke P (2023) Ecotoxicity characterization of chemicals: global recommendations and implementation in USEtox. Chemosphere 310:136807. https://doi.org/10.1016/j.chemosphere.2022.136807

Payet J (2004) Assessing toxic impacts on aquatic ecosystems in life cycle assessment (LCA) Verlag nicht ermittelbar].

Roos S, Holmquist H, Jönsson C, Arvidsson R (2017) USEtox characterisation factors for textile chemicals based on a transparent data source selection strategy [OriginalPaper]. The International Journal of Life Cycle Assessment 23(4):890–903. https://doi.org/10.1007/s11367-017-1330-y

Rosenbaum RK, Bachmann TM, Gold LS, Huijbregts MA, Jolliet O, Juraske R, Koehler A, Larsen HF, MacLeod M, Margni M (2008) USEtox—the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. Int J Life Cycle Assess 13:532

Rosenbaum RK (2015) Ecotoxicity. In M. Z. Hauschild & M. A. J. Huijbregts (Eds.), Life cycle impact assessment (pp. 139–162). Springer Netherlands. https://doi.org/10.1007/978-94-017-9744-3_8

Sahu JK, Ganguly S, Kaushik A (2013) Triazoles: a valuable insight into recent developments and biological activities. Chin J Nat Med 11(5):456–465. https://doi.org/10.1016/S1875-5364(13)60084-9

Schenker U, Scheringer M, Hungerbühler K (2007) Including degradation products of persistent organic pollutants in a global multimedia box model. Environmental Science and Pollution Research - International 14(3):145–152. https://doi.org/10.1065/espr2007.03.398

Schulze C, Jödicke A, Scheringer M, Margni M, Jolliet O, Hungerbühler K, Matthies M (2001) Comparison of different life-cycle impact assessment methods for aquatic ecotoxicity. Environmental Toxicology and Chemistry: an International Journal 20(9):2122–2132

U.S.EPA (2020) User's guide for T.E.S.T. (version 5.1) (toxicity estimation software tool): a program to estimate toxicity from molecular structure [Data and Tools]. https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

U.S.EPA (2023a) ECOTOX | explore. Retrieved September from https://cfpub.epa.gov/ecotox/explore.cfm?sub=Species

U.S.EPA (2023b) Estimation Programs InterfaceSuite™ for Microsoft® Windows, v 4.11. Retrieved March from https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface

van Zelm R, Huijbregts MAJ, Posthuma L, Wintersen A, van de Meent D (2009) Pesticide ecotoxicological effect factors and their uncertainties for freshwater ecosystems. The International Journal of Life Cycle Assessment 14(1):43–51. https://doi.org/10.1007/s11367-008-0037-5

Van Zelm R, Huijbregts MA, van de Meent D (2010) Transformation products in the life cycle impact assessment of chemicals. Environ Sci Technol 44:1004

VEGA HUB (2022) Retrieved August from https://www.vegahub.eu/about-vegahub/

Wicker J, Lorsbach T, Gütlein M, Schmid E, Latino D, Kramer S, Fenner K (2016) enviPath–the environmental contaminant biotransformation pathway resource. Nucleic Acids Res 44(D1):D502–D508

Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, Richard AM (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminformatics 9(1):61. https://doi.org/10.1186/s13321-017-0247-6

Wolfe K, Pope N, Parmar R, Galvin M, Stevens C, Weber E, Flaishans J, Purucker T (2016) Chemical transformation system: cloud based cheminformatic services to support integrated environmental modeling.

Yuan C, Tebes-Stevens C, Weber EJ (2021) Prioritizing direct photolysis products predicted by the chemical transformation simulator: relative reasoning and absolute ranking. Environ Sci Technol 55(9):5950–5958. https://doi.org/10.1021/acs.est.0c08745

Zelm R, Huijbregts M, Harbers J, Wintersen A, Struijs J, Posthuma L, Van de meent D (2007) Uncertainty in msPAF-based ecotoxicological effect factors for freshwater ecosystems in life cycle impact assessment. Integr Environ Assess Manag 3:203–210. https://doi.org/10.1002/ieam.5630030422