**LCIA OF IMPACTS ON HUMAN HEALTH AND ECOSYSTEMS**

CrossMark

# The potential to use QSAR to populate ecotoxicity characterisation factors for simplified LCIA and chemical prioritisation

Hanna Holmquist[1] · Jenny Lexén[2] · Magnus Rahmberg[2] · Ullrika Sahlin[3] · Julia Grönholdt Palm[3] · Tomas Rydberg[2]

## Abstract

**Purpose** Today's chemical society use and emit an enormous number of different, potentially ecotoxic, chemicals to the environment. The vast majority of substances do not have characterisation factors describing their ecotoxicity potential. A first stage, high throughput, screening tool is needed for prioritisation of which substances need further measures.

**Methods** USEtox characterisation factors were calculated in this work based on data generated by quantitative structure-activity relationship (QSAR) models to expand substance coverage where characterisation factors were missing. Existing QSAR models for physico-chemical data and ecotoxicity were used, and to further fill data gaps, an algae QSAR model was developed. The existing USEtox characterisation factors were used as reference to evaluate the impact from the use of QSARs to generate input data to USEtox, with focus on ecotoxicity data. An inventory of chemicals that make up the Swedish societal stock of plastic additives, and their associated predicted emissions, was used as a case study to rank chemicals according to their ecotoxicity potential.

**Results and discussion** For the 210 chemicals in the inventory, only 41 had characterisation factors in the USEtox database. With the use of QSAR generated substance data, an additional 89 characterisation factors could be calculated, substantially improving substance coverage in the ranking. The choice of QSAR model was shown to be important for the reliability of the results, but also with the best correlated model results, the discrepancies between characterisation factors based on estimated data and experimental data were very large.

**Conclusions** The use of QSAR estimated data as basis for calculation of characterisation factors, and the further use of those factors for ranking based on ecotoxicity potential, was assessed as a feasible way to gather substance data for large datasets. However, further research and development of the guidance on how to make use of estimated data is needed to achieve improvement of the accuracy of the results.

**Keywords** Characterisation factors · Plastic additives · Prioritisation · QSAR · USEtox

✉ Hanna Holmquist
  hanna.holmquist@chalmers.se

[1] Division of Environmental Systems Analysis, Chalmers University of Technology, 412 96 Gothenburg, Sweden

[2] IVL Swedish Environmental Research Institute Ltd., 400 14 Gothenburg, Sweden

[3] Lund University Centre for Environmental and Climate Research, 223 62 Lund, Sweden

## 1 Introduction

Every day a wide variety of chemicals are emitted into the environment from a multitude of sources. These emissions and the subsequent pollution of the natural environment and potential exposure of living organisms and humans may pose a risk to the ecosystem and human health (UNEP 2012). To efficiently reduce this risk by implementing reduction measures or substitution, it is necessary to identify chemical emissions of concern (e.g. Egeghy et al. 2011; von der Ohe et al. 2011). The reason for concern should be based on the potential negative effect of the chemicals rather than only emitted amount. Chemical risk assessment is one way to obtain

the relevant information (e.g. van Leeuwen and Vermeire 2007), but that is generally a very time-consuming process. There is a need for a fast and easy-to-use screening tool, based on (eco)toxicity potential but not necessarily a full risk assessment, to be able to do a first prioritisation for large datasets.

The USEtox model can be used to integrate data on chemical fate and effect into measures of potential impact and should thus have a good potential to be used for prioritisation of chemicals. USEtox is a scientifically based, consensus model to characterise potential impacts on human health and freshwater aquatic ecosystems from a product (or service) life cycle (Hauschild et al. 2008; Rosenbaum et al. 2008). With the USEtox model, a large number of characterisation factors (CFs) have been published but the model can also be used to calculate new CFs. To run the USEtox model is a fast process, but the collection of necessary substance data, in particular the (eco)toxicological data, can be time-consuming, especially if the chemical inventory is large.

The use of (quantitative) structure-activity relationships (QSAR) models for the estimation of the properties of chemicals for which data are needed can be one way to speed up the screening process, especially for cases where experimental data are lacking. A QSAR model is a relation between chemical structure and a property of the chemical compound (ECHA 2016; OECD 2007b). The features of a chemical structure are captured by a set of chemical descriptors that are used to predict characteristics of the chemical. In comparison to other data collection methods, QSARs are fast and have the potential to cover many substances. The substance coverage is dependent of the model domain and, in contrast to, e.g. Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) registration data, independent of the substance manufacturing volumes. In addition, estimated data, e.g. by QSARs, are likely to increase in importance as authorities such as the European Chemicals Agency promote a reduced use of animal testing (ECHA 2017).

The aim of the present study was to test if the use of substance data predicted by QSARs can be one way to use the USEtox model as part of a fast and easy screening tool, with broad coverage, for ranking within large datasets based on ecotoxicity potential. For this purpose, an emission inventory for a wide range of plastic additives, such as pigments, flame retardants, stabilisers and plasticisers, was used as a case study. The focus of the present study was on ecosystem effects, and thus, toxicity data on human health were not collected or predicted.

## 2 Methods

### 2.1 The USEtox model

With the USEtox model, (eco)toxicity CFs integrate the fate, exposure and effects of a chemical after emission into the environment. Emission compartment specific CFs are calculated from the product of matrices containing fate factors (FF), freshwater ecosystem exposure factors (XF) and freshwater aquatic ecosystem toxicity effect factors (EF) (Eq. (1)). The CFs provide the means to convert the chemical emissions into impact scores (IS) and thus compare chemical emissions based on (eco)toxicity potential.

$$CF = EF \ XF \ FF \qquad (1)$$

Each of the factors (FF, XF and EF) is calculated with the use of substance specific data on physico-chemical properties as well as (eco)toxicological effects. The obligatory parameters for organic substances, necessary to provide the model with, are listed in Table 1. Remaining parameters can be estimated with model internal estimation routines, as long as the substance is within the model domain.

USEtox CFs are defined as recommended if the EF is based on data from three trophic levels and indicative if the ecotoxicity data cover less than three different trophic levels (Huijbregts et al. 2015b). The standard ecotoxicological test set for aquatic organisms covers primary producers, invertebrates and fish, often as algae, the water flea *Daphnia magna* and fish such as rainbow trout or fathead minnow.

To calculate CFs in this study, the USEtox model version 2.02 was applied. The focus of the present study was on ecosystem effects and thus only ecotoxicity CFs were calculated. The model user manual (Huijbregts et al. 2015a) and the manual for the organic substance database (Huijbregts et al. 2015b) were used as basis for the workflow and data

**Table 1** Parameters required for organic substances in USEtox 2.02

| Parameter | Symbol | Unit |
| --- | --- | --- |
| Molar mass | MW | g/mol |
| p$K$a chemical class | pKaChemClass | – |
| p$K$a base reaction | pKa.gain | – |
| p$K$a acid reaction | pKa.loss | – |
| Partitioning coefficient between *n*-octanol and water | $K_{OW}$ | L/L |
| Vapour pressure (at 25 °C) | Pvap25 | Pa |
| Solubility (at 25 °C) | Sol25 | mg/L |
| Rate constant degradation in air | kdegA | 1/s |
| Rate constant degradation in water | kdegW | 1/s |
| Rate constant degradation in sediment | kdegSd | 1/s |
| Rate constant degradation in soil | kdegSl | 1/s |

Source: Huijbregts et al. 2015a; Table 2

collection. The USEtox manuals (Huijbregts et al. 2015a; Huijbregts et al. 2015b) recommend the use of experimental data but provide guidance also to the use of estimated data for the physico-chemical data collection. Indeed, many of the existing USEtox CFs have fate factors based on estimated data. The USEtox manuals do however not give any further guidance on the estimation of ecotoxicological effects, and thus, the identification of available models, suitable for the purpose, was part of the present study.

## 2.2 Chemical inventory

The research programme ChEmiTecs has developed a simple method for an initial approximation of emissions of organic chemicals from products containing plastic materials (Bilitewski et al. 2012). The method has been used to estimate annual emissions of organic chemicals from the accumulated stock of products containing plastic materials in the Swedish society. The output from this modelling approach was rough emission estimates for 210 specified organic substances, and those were used as our case study inventory. More recent estimates based on a rough extrapolation from a much smaller sample of parallel calculations with a more sophisticated model indicate that the emissions in the first inventory were overestimated by, on average, a factor of in the order of 100 but with very large individual variations among substances and materials (Palm Cousins et al. 2018; Rydberg and Lexén 2016). In the study presented here, however, the original plastic additives inventory serves well as a case study to test the usability of QSAR-based CFs (the inventory is included in the Electronic Supplementary Material). The absolute magnitude of emissions is not the key feature of this paper but that these plastic additives have varying physico-chemical and ecotoxicological properties, which is why it is of interest not only to assess the emission loads but also the chemicals' potential to exert negative impacts on the environment (and humans though that was not part of the present study). For the majority of these substances, 169 of 210, USEtox characterisation factors (CFs) were not available.

## 2.3 Estimation of substance data with QSARs

The 210 substances in the plastic additive inventory were each identified by name, chemical abstract service (CAS) number and simplified molecular-input line-entry system (SMILES) codes. QSAR models were run, aiming to generate the relevant physico-chemical properties as well as ecotoxicological data for CF calculation with USEtox for all 210 substances. USEtox manuals give detailed guidance on how to obtain physico-chemical data by estimation methods if experimental data are lacking. Therefore, focus on model selection evaluation was placed on the ecotoxicity data, for which detailed guidance is lacking for generation of estimated data.

Existing QSAR models were used and selected based on the relevance for the purpose, i.e. whether the generated data were relevant as input data to USEtox. The existing QSAR models have already been validated, and model validation was not part of the scope of the present study. It should be noted that a QSAR model is only valid in its applicability domain, and any model used should specify if the prediction of a chemical property is in or out of the domain. The SMILES codes were used as input to the QSAR models.

### 2.3.1 Physico-chemical properties

Physical chemical properties were predicted with the U.S. Environmental Protection Agency Estimation Program Interface (EPI) Suite, version 4.11 (US EPA 2012). The EPI Suite model is recommended for use in the USEtox manuals and the guidelines therein were followed. Substance data for CF calculation were only collected or estimated for the parameters listed as necessary in the USEtox manual (Huijbregts et al. 2015a), see Table 1. The application domain of EPI Suite is organic chemicals, and inorganic as well as organo-metallic chemicals are outside the model domain. The plastic additives chemical inventory contained some organometallic substances which were thus outside the model domain, and physico-chemical properties obtained were not assessed as valid, but not removed from the dataset. Whether the EPI Suite training sets contain substances similar to the plastic additives in the case study inventory was not controlled as it was considered to be out of scope for the fast screening aimed for here. Although the USEtox manual states that preference should be given to experimental values only estimated values were used. The reason was that differences between estimated and experimental values for the water solubility were very large for some substances, and we reasoned that using only the estimated values (for all parameters), we would achieve a better comparability between results within the dataset. These differences can be due to measurement errors or variability in experimental data or predictive errors in the QSAR estimates. The SPARC software is recommended in the USEtox manual for prediction of acid and base dissociation constants ($pK_a$/$pK_b$) (Huijbregts et al. 2015b), but this is a commercial software available at a fee. Acid and base dissociation constants were not estimated, as we are here focussing on the ecotoxicological data, and all substances were entered into the USEtox model as neutral substances, with pKa.gain 0 and pKa.loss 14. The ignorance of dissociation properties adds uncertainty to the results, especially for anionic substances (Rybacka and Andersson 2016). This is however in line with previous versions of USEtox where dissociating substances were not modelled differently to neutral substances but flagged as interim. The physico-chemical data used to calculate the CFs are included in the Electronic Supplementary Material.

### 2.3.2 Aquatic ecotoxicity data

There are several guidelines available for the use of QSAR. These guidelines describe how to use and report QSAR. They all are in various aspects based on the OECD principles of QSAR. These principles were agreed on in 2004 and published in the Guidance Document on the Validation of (Quantitative) Structure Activity Relationship [(Q)SAR] Models (OECD 2007a). The OECD principles states: to facilitate the consideration of a QSAR model for regulatory purposes, the model should be associated with the following information:

1. A defined endpoint
2. An unambiguous algorithm
3. A defined domain of applicability
4. Appropriate measures of goodness-of-fit, robustness and predictivity
5. A mechanistic interpretation, if possible

The criteria listed above were applied when choosing QSAR models to be used for calculation of the aquatic ecotoxicity.

Since EPI Suite is recommended for the estimation of physico-chemical data, its Ecological Structure Activity Relationships (ECOSAR) Predictive Model was a natural starting point to generate aquatic ecotoxicity data. ECOSAR model v.1.11, included in the EPI Suite package, was applied for the purpose. The ECOSAR model domain cover organic chemicals and inorganic chemicals, organometallic chemicals and polymers are out of domain. The ECOSAR model domain has further restrictions based on the limitations of the training set (not investigated further herein), and the model gives results also for substances that are not within the model domain, thus potentially introducing large uncertainties. ECOSAR was run for the full dataset, including the organometallics out of model domain. The endpoints estimated were fish $LC_{50}$ (concentration where 50% of the population exhibit a response) 96 h, daphnia $EC_{50}$ 48 h and green algae $EC_{50}$ 72 h. When more than one datum was estimated for each endpoint, the lowest datum was selected for the further modelling since false positives (i.e. overestimated ecotoxicity) are preferable to false negatives (underestimated ecotoxicity) in a screening context.

Given the uncertainty introduced with ECOSAR, aquatic ecotoxicity data were also estimated with the Toxicity Estimation Software Tool (TEST), version 4.2.1, also from US Environmental Protection Agency (US EPA 2016). The endpoints estimated were fathead minnow $LC_{50}$ 96 h and *Daphnia magna* $LC_{50}$ 48 h. They were estimated by the consensus model in TEST as it has been shown that consensus models give a better estimate of the toxicity and also covers a larger applicable domain than the individual models (Zhu et al. 2008). The consensus model is the average value from the predicted values for several different methods. The result from the consensus model is only based on the methods which have predicted results in the applicability domain of the corresponding model, i.e. the uncertainty from using results for substances outside the model domain was removed.

A QSAR for algae toxicity (endpoint $EC_{50}$ *Pseudokirchneriella subcapitata* (72–96 h)) was developed to complement the TEST model with a third phylum. The QSAR model was built using partial least square (PLS) regression with molecular descriptors calculated from the Dragon 6.0 software (TALETE 2014) and experimental algal toxicity data on 80 substances (Grönholdt Palm 2014). The model has a defined applicability domain and model performance was evaluated as satisfying with respect to training ($n = 48$, $R^2 = 0.86$, RMSEE = 0.53) and testing ($n = 22$, Q2CV = 0.72, RMSEP = 0.97).

The above described procedure generated three aquatic ecotoxicity datasets, listed in Table 2. These datasets were used for calculation of QSAR-based CFs with USEtox. The estimated values for E/$LC_{50}$ were used for calculation of the concentration where 50% of species are exposed above their E/$LC_{50}$ (i.e. the $HC_{50}$), which is needed in USEtox for the calculation of the ecotoxicological effect factor (EF). Chronic $HC_{50}$ values are preferable, but in this case, only acute data were estimated. Instead, the chronic-equivalent was derived by dividing the acute values by 2 (the acute-to-chronic extrapolation factor) as described in Huijbregts et al. (2015a). The EFs are listed in the Electronic Supplementary Material.

Other ecotoxicity QSAR models and data estimation methods are available, e.g. via the OECD QSAR toolbox (OECD 2011), but not used herein. The ECOSAR QSARs are incorporated into the OECD QSAR Toolbox together with additional QSARs for fathead minnow (*Pimephales promelas*). The OECD QSAR Toolbox contains other functionalities for data gap filling, in addition to QSARs, but the automated workflows are for fish endpoints only (QSAR Toolbox version 4.1) and the categorisation procedure for read-across is not

**Table 2** Datasets used for calculation of USEtox characterisation factors

| Dataset | Data generator | |
|---|---|---|
| | Physico chemical data | Ecotoxicological data |
| 1 | EPI Suite | ECOSAR |
| 2 | EPI Suite | TEST |
| 3 | EPI Suite | TEST + algae model |

simple, and the result depends on the experience and skills of the modeller (Dimitrov et al. 2016).

## 2.4 Assessing the relevance of QSAR-based CFs

The reliability of the QSAR-based CFs was evaluated by the difference and correlation to the USEtox CFs based on the organic substances database of USEtox 2.02. Possible differences were further evaluated by studying correlations on the factors FF, IF and EF and by between substance comparisons. The substances in the inventory that had USEtox CFs were not organometallics or polymers and were thus within the broad model domain definitions as described above. The strength of the linear relationship between the USEtox CFs and the QSAR-based CFs (datasets 1–3; Table 2) was evaluated by comparing the squared correlation coefficients ($R^2$) from regression analysis where the QSAR-based CFs were regressed on the USEtox CFs. Pairwise comparisons of the differences between the QSAR-based CFs and the USEtox CFs were made to evaluate which QSAR-based CFs that deviated the least from the USEtox CFs. To explore the impact of the modelling choices in the parameterisation for the FF, i.e. the inclusion of only estimated data and the ignorance of dissociation constants (see Sect. 2.3.1), CFs based on physico-chemical data with preference to experimental data as recommended in the USEtox manual (Huijbregts et al. 2015b) were calculated for the limited dataset that already had USEtox CFs (dataset 4; data listed in the Electronic Supplementary Material). For this limited dataset, to further explore the impacts of the deviations from the USEtox manuals, dissociation constants were estimated by use of Marvin Sketch 17.24 (ChemAxon 2017) and included in the input data (substances with $pKa > 14$ or $< 0$ were included as neutrals).

## 2.5 Calculation of impact score and ranking of chemicals

IS (Eq. (2)) for the plastics additives were calculated according to the principles in the USEtox manual (Huijbregts et al. 2015a). The emission inventory was based on diffuse emissions from plastic products and since majority of the products will probably be in air, 90% of emissions were arbitrarily assigned to urban air and the remaining 10% of emission to continental freshwater.

$$IS_{aq.ecotox.} = CF_{f.w.} \times M_{f.w.} + CF_{u.a.} \times M_{u.a.} \qquad (2)$$

(IS = impact score in comparative toxic units ($CTU_e$), f.w. = freshwater, u.a. = urban air, $M$ = emitted mass)

The IS was used to rank the plastic additive emissions according to their relative aquatic ecotoxicity potential. The plastic additives were grouped into functional categories for comparison of functional group ranking based on emitted mass and ecotoxicity potential.

# 3 Results

## 3.1 Assessing the relevance of QSAR-based CFs

Of the new QSAR-based CFs, there was an overlap with the 41 USEtox CFs for 38 and 35 substances, respectively, using the ECOSAR model (dataset 1) and TEST model (dataset 2) as ecotoxicity data generator. The 35 substances in dataset 2 were also included in dataset 1, and the ECOSAR model was able to generate ecotoxicity data for three additional substances. With the addition of algae data from the algae model (dataset 3), the dataset of overlapping CFs was reduced to only 10 substances. These overlapping CF datasets were used for the further analysis to assess the relevance of the QSAR-based CFs. The data were strongly skewed and were $log_{10}$-transformed to adjust this skewedness. No apparent outliers were noticeable in the log-transformed datasets, and all data were included in the analysis.

Figure 1a shows the correlation between the ECOSAR based CFs and the USEtox CFs, which has an evident large variation; the explained variance ($R^2$) for the regression of the log-transformed data ranged from 15 to 51%, depending on emission compartment. Figure 1b shows that the TEST based CFs, based on ecotoxicity data for only fish and daphnia, were better correlated to the USEtox CFs; $R^2$ for the regression of the log-transformed data ranged from 78 to 86%, depending on emission compartment. The improvement in CF correlation when the TEST model was used to derive EF instead of the ECOSAR model cannot be explained by the exclusion of substances out of model domain as there was almost a complete overlap in the datasets. With the addition of algae data, the dataset was reduced to only 10 substances, which limit the possibilities for conclusions. Also for this small dataset, correlations between CFs were low; $R^2$ for the regression of the $log_{10}$-transformed data ranged from 48 to 85%, depending on emission compartment. Residual errors (RE) and parameter estimates from the linear regression model of log-transformed CFs based on QSAR data (ecotoxicological data from the TEST model) by the log-transformed CFs from USEtox are listed in Table 3. The constant for the slope of the line ($\beta_1$) was not different from 1 ($p < 0.05$), and the intercept was between 0.51 and 0.81, depending on emission compartment. Since the slope does not significantly differ from 1, the untransformed relationship between the CFs based on estimated data and the USEtox CFs is equal to $10^{\beta_0}$, approximately 3-6 and thus on average predicting a higher ecotoxicity potential than the USEtox CFs. The 95% confidence intervals
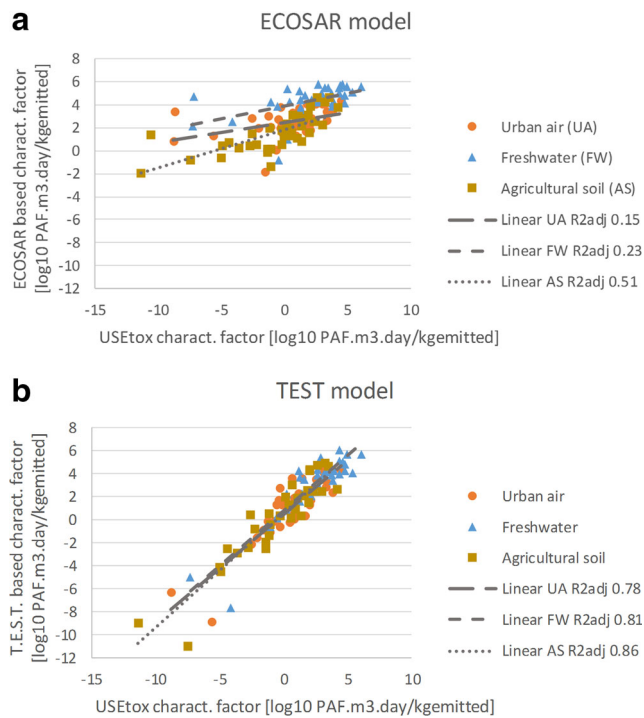
**Fig. 1** Correlation between characterisation factors based on QSAR-data (**a** ECOSAR model, **b** TEST model) and USEtox characterisation factors. Note the logarithmic scale

for the regression (± 2 RE), equal multiplication, respectively division, with $(10^{RE})^2$ for the untransformed relationship: 300–500 CTU$_e$, depending on emission compartment.

To further illustrate the differences in CFs based on estimated data and USEtox CFs, Fig. 2 shows boxplots for the differences between the QSAR-based CFs and USEtox CFs. The spread around zero, i.e. no deviation between USEtox CF and the CF based on estimated data, is smaller

**Table 3** Residual error (RE) and parameter estimates for the regression models where log10-transformed CFs based on estimated data (TEST model used for ecotoxicity) were regressed on log10-transformed CFs from USEtox, for emissions to urban air, freshwater and agricultural soil

| Term | | Estimate | Lower 95% | Upper 95% |
|---|---|---|---|---|
| **Emission to urban air** | | | | |
| RE | 1.32 | | | |
| Intercept, $\beta_0$ | | 0.51 | 0.049 | 0.98 |
| Constant, $\beta_1$ | | 0.94 | 0.76 | 1.1 |
| **Emission to freshwater** | | | | |
| RE | 1.24 | | | |
| Intercept, $\beta_0$ | | 0.81 | 0.25 | 1.4 |
| Constant, $\beta_1$ | | 0.95 | 0.79 | 1.1 |
| **Emission to agricultural soil** | | | | |
| RE | 1.35 | | | |
| Intercept, $\beta_0$ | | 0.66 | 0.19 | 1.1 |
| Constant, $\beta_1$ | | 0.99 | 0.86 | 1.1 |

for the TEST based CFs compared to those based on ECOSAR data. Still, the deviations are large for some substances, especially for the freshwater compartment, and the median deviation for CFs based on ecotoxicological data generated by TEST was 0.1, 350 and 0.31 for emissions to urban air, continental freshwater and agricultural soil, respectively.

To try and explain the variance observed between CFs from USEtox and the QSAR-based CFs, FF, XF and EFs were compared by regression of the QSAR-based factors on the USEtox factors. There was a large variation in the FFs and fitting a line to the log-transformed data, the adjusted $R^2$ was 39 and 24% for the FF for freshwater and emission to urban air and freshwater, respectively. These differences apply for all the QSAR-based CFs as the same data for physico-chemical parameters were used. Data for the obligatory input parameters (Table 1) that deviated the most from USEtox substance data were data for water solubility and chemical class classification (acid/base/amphoteric/neutral). The median difference in FF between the QSAR-based and USEtox FFs was − 0.17 and − 0.045 days for the freshwater compartment and emissions to freshwater and urban air, respectively. The XFs showed a good correlation and the adjusted $R^2$ was 93% for the untransformed data. Three highly lipophilic substances could be identified as possible outliers and those were predicted to be less bioavailable by the QSAR-based approach compared to the USEtox XF. The correlation between the EF based on TEST-generated data and USEtox EF was slightly better compared to the FFs; with the adjusted $R^2$ at 43% for the log-transformed EFs. The median difference in EF was however as large as 1295 PAF × m$^3$/kg between TEST and USEtox and 16,288 PAF × m$^3$/kg for ECOSAR and USEtox. Since the deviation was smaller, the QSAR-based CFs using the TEST model to estimate ecotoxicity data generated the CFs most similar to the USEtox CFs in this case. The distribution of deviations also shows that the EF was the more influential factor between the EF and the FF as the deviation was larger and the CF is directly proportional these factors.

Giving priority to experimental data, and accounting for dissociation (additional CFs calculated for a the limited dataset; see Sect. 2.3.1), improved the CFs similarity with USEtox CFs as this procedure was in line with the USEtox procedures, but the improvement was minor. Correlation analysis between the USEtox CFs and the CFs calculated with QSAR data for ecotoxicity (generated by the TEST model) and experimental data when available, but otherwise estimated data for physico-chemical properties, showed an explained variance between 80 and 85% (cf. $R^2$ 78–85% when only estimated data were used). For this correlation analysis, one substance (CAS 6683-19-8, tetrakis methylene(3,5-di-t-butyl-4-hydroxyhydrocinnamate)methan) was removed from the dataset since it was considered an outlier due to the large deviation between the estimated $K_{OW}$ and the experimental $K_{OW}$.

**Fig. 2** Boxplots for the pairwise differences between and the characterisation factors based on QSAR-data and the USEtox characterisation factors
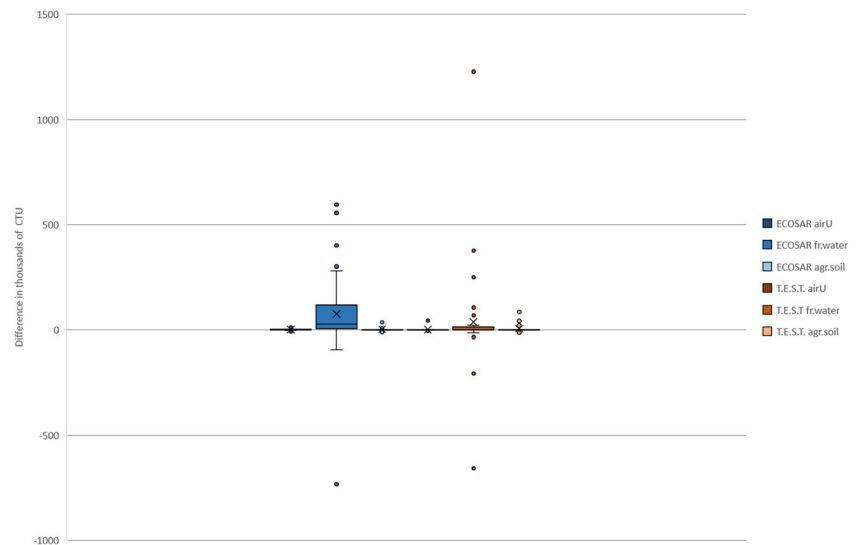
## 3.2 New characterisation factors for plastic additives

The use of QSAR generated data improved the coverage of CFs for the plastic additives chemical inventory greatly, going from 41 to 170, or 124, depending on model selection (Table 4). The algae model developed within the present study did not cover more than 28 of the substances in the inventory but can be used together with USEtox CFs to improve coverage if CFs based on three trophic levels are needed. All CFs calculated within the present study are made available in the Electronic Supplementary Material.

## 3.3 Uncertainty and precision

The precision, with regard to model uncertainty, of the USEtox CFs is within a factor of 10–100 for freshwater ecotoxicity, and this needs to be considered when assessing contributions to the total toxicity score (Rosenbaum et al. 2008). Eleven to 15 substances (depending on emission compartment), out of the 210, contributed significantly to the total sum of CF, considering the model uncertainty (substances with a CF that contribute to more than 1% of the sum of CFs).

**Table 4** Characterisation factors available or possible to calculate with QSAR-data, for the inventory of 210 plastic additive chemicals

| USEtox | ECOSAR | TEST | TEST + algae model | TEST + algae model + USEtox |
|--------|--------|------|--------------------|-----------------------------|
| 41 | 170 | 124 | 28 | 59 |

In addition to the model uncertainty, there is also parameter uncertainty. Parameter uncertainty exists also in the USEtox CFs and is likely increased in the QSAR-based CFs calculated within the present study. To quantify the additional uncertainty, introduced by the use of estimated data instead of experimental data, the statistical procedure that Rosenbaum et al. (2008) applied in the model comparison, based on McKone (1993) was applied also here. A 95% confidence interval was used to generate a quantification of the orders of magnitude of the added uncertainty. The 95% confidence interval for the regression estimate was division/multiplication with 300–500 $CTU_e$ and, in analogy to Rosenbaum et al. (2008), the factor describing the additional uncertainty approximately 100–1000. Adding this uncertainty to the model uncertainty makes the total uncertainty three to five orders of magnitude and thus impractical for implementation. In this case, with the plastic additives, the CFs range over 18 orders of magnitude or more and it would be possible to differentiate between substances contributing to the risk score and those with no significant contribution. But the large bulk of substances would still have an unknown contribution score if uncertainties were accounted for.

## 3.4 Prioritisation and ranking of chemical emissions

The TEST-based CFs (dataset 2) were used to generate IS, since those were the QSAR-based CFs most similar to the USEtox CFs. The IS integrate the emitted amount from the emissions inventory with the fate, exposure and effect as quantified in the CF. Taking the above-mentioned uncertainty into account, 37-86 substances could be identified as the ones

with the highest ecotoxicity potential as those were within three to five orders of magnitude of the total IS. The substances giving the largest contributions to the overall ecotoxicity potential score belong to several different functionality categories, e.g. pigments, flame retardants, stabilisers, plasticisers and lubricants. Flame retardants and plasticisers were estimated to be emitted in the largest amounts contributing to 36 and 28% of total emissions, respectively, with the remaining additive categories contributing with less than 10% each to the total emitted amount. The IS on the other hand imply that also pigments and stabilisers (bio) impose a risk as also they contribute with more than 10% to the total IS.

## 4 Discussion and conclusions

The main aim of this study was to assess the possibilities to use QSAR-based data in the LCIA model USEtox to rank chemical emissions according to the predicted relative aquatic ecotoxicity potential. The results show that QSAR data can indeed be used for a fast calculation of CFs with the USEtox model and that those CFs could be used to make prioritisation in large inventories as substance coverage can be markedly increased compared to if only existing USEtox CFs are used. However, it was also shown that substantial uncertainty was added to the CFs, limiting the current practical use. The choice of QSAR model was shown to be crucial for the relevance and robustness of the outcome and in the present study two ecotoxicity estimation models were used, and the TEST model was shown to provide the best estimations when data were to be used to calculate CFs as similar as possible to USEtox CFs, despite that it only cover species from two trophic levels.

Alfonsín et al. (2014), Igos et al. (2014) and Roos et al. (2017) could show that the effect data is by far the most influential parameter when CFs are calculated with the USEtox model, and our results are in line with this finding as we could see that the variation in the EF was the most important contributor to the variation seen in the CF in the comparison between the QSAR-based CFs and the USEtox CFs. Interestingly, we could also see that the FF can be another important contributor to the CF variation, since there was a large difference in property data for some substances and parameters. Recently, Roos et al. (2017) proposed the increased use of QSAR estimated data to calculate CFs as a way to fill data gaps where experimental data are missing. Roos et al. (2017) assigned the CF qualitative uncertainty scores indicating high uncertainty for CFs calculated based on estimated data alone. The present study, with a large QSAR-based dataset, quantitatively shows how large the uncertainties can be as the additional uncertainty from the use of QSARs to generate input data was quantified to a factor 100–1000. In the present study, the USEtox CFs, with FF based on estimated and experimental data, and EF mainly

based on experimental data, were considered the baseline, not considering uncertainties in that dataset. The uncertainty quantified herein could therefore be under- or overestimated in relation to the true uncertainty. In any case, since the added uncertainty is large, further studies investigating how QSAR models can be used for calculation of CFs are needed. This includes population also of the input data for calculation of the FF and XF, e.g. dissociation constants, but mainly an expanded overview and comparison between available and relevant models to generate ecotoxicity data.

It can be argued that experimental data made available under the REACH legislation should be the first-hand choice for population of the USEtox model input data. Several authors have indeed tested the usability of European data sources, e.g. Müller et al. (2017), Saouter et al. (2017a, b), and have concluded that the USEtox model and procedures need to be adapted to make it possible to use all available data, e.g. by allowing the use of chronic data expressed in other forms than $EC_{50}$. Despite the huge increase in data availability for experimental data on physico-chemicals properties as well as (eco)toxicity that comes with the REACH registrations, a parallel line of work on the inclusion of estimated (QSAR) data in CF calculation is also warranted, as QSAR data will probably be an important data source in the future in any case, as the reduction of animal testing is one of the EU goals (ECHA 2017).

In addition to the model and parameter uncertainty, there is also uncertainty added from the limitations of model scope, e.g. that USEtox only includes freshwater ecotoxicity and thus excludes effects in the marine or benthic compartments. Other models are available and depending on model structure and scope the results can differ, for example, Mattila et al. (2011) compared three models and identified USEtox to give a lot of focus to metals, and thus, metal contribution will be given a high weight using USEtox in cases where the inventory contain metals.

To conclude, the use of QSAR models to generate data for calculation of CFs has potential to fill data gaps and allow for a first screening of large inventories based on ecotoxicity potential. The uncertainty added by using estimated data is however a limiting factor, and further research is needed to develop recommendations on what models to use and possibly also to develop better models.

# References

Alfonsín C, Hospido A, Omil F, Moreira MT, Feijoo G (2014) PPCPs in wastewater—update and calculation of characterization factors for their inclusion in LCA studies. J Clean Prod 83:245–255

Biletewski B et al (eds) (2012) Emissions of additives from plastics in the societal material stock: a case study for Sweden. In: Global risk-based management of chemical additives I. Springer, pp 253–264

ChemAxon (2017) Marvin Sketch 17.24.0. www.chemaxon.com. Accessed 2018–01-19

Dimitrov SD, Diderich R, Sobanski T, Pavlov TS, Chankov GV, Chapkanov AS, Karakolev YH, Temelkov SG, Vasilev RA, Gerova KD, Kuseva CD, Todorova ND, Mehmed AM, Rasenberg M, Mekenyan OG (2016) QSAR toolbox—workflow and major functionalities. SAR QSAR Environ Res 27:203–219. https://doi.org/10.1080/1062936X.2015.1136680

ECHA (2016) Practical guide how to use and report (Q)SARs. Version 3.1. doi:https://doi.org/10.2823/81818

ECHA (2017) What about animal testing? https://echa.europa.eu/chemicals-in-our-life/animal-testing-under-reach. Accessed 2017–09-25

Egeghy PP, Vallero DA, Cohen Hubal EA (2011) Exposure-based prioritization of chemicals for risk assessment. Environ Sci Pol 14:950–964. https://doi.org/10.1016/j.envsci.2011.07.010

Grönholdt Palm J (2014) The impact of algal toxicity on life-cycle impact assessment of plastic additives and the potential of using QSAR predictions to fill the algae data gap. Master thesis 30 ECTS Lund University. http://lup.lub.lu.se/student-papers/record/5364266

Hauschild MZ, Huijbregts M, Jolliet O, Macleod M, Margni M, van de Meent D, Rosenbaum RK, McKone TE (2008) Building a model based on scientific consensus for life cycle impact assessment of chemicals: the search for harmony and parsimony. Environ Sci Technol 42:7032–7037

Huijbregts M, Margni M, Hauschild M, Jolliet O, McKone T, Rosenbaum R, Meent DVD (2015a) USEtox® 2.0 user manual (version 2)

Huijbregts M, Meent DVD, Margni M, Jolliet O, Rosenbaum R, McKone T, Hauschild M (2015b) USEtox 2.0 manual: organic substances (version 2)

Igos E, Moeller R, Benetto E, Biwer A, Guiton M, Dieumegard P (2014) Development of USEtox characterisation factors for dishwasher detergents using data made available under REACH. Chemosphere 100:160–166. https://doi.org/10.1016/j.chemosphere.2013.11.041

Mattila T, Verta M, Seppälä J (2011) Comparing priority setting in integrated hazardous substance assessment and in life cycle impact assessment. Int J Life Cycle Assess 16:788–794

McKone TE (1993) The precision of QSAR methods for estimating intermedia transfer factors in exposure assessments. SAR QSAR Environ Res 1:41–51

Müller N, de Zwart D, Hauschild M, Kijko G, Fantke P (2017) Exploring REACH as a potential data source for characterizing ecotoxicity in life cycle assessment. Environ Toxicol Chem 36:492–500

OECD (2007a) Guidance document on the validation of (quantitative) structure activity relationship [(Q)SAR] models. OECD series on testing and assessment no. 69. ENV/JM/MONO(2007)2

OECD (2007b) Guidance on grouping of chemicals. Series on testing and assessment number 80. ENV/JM/MONO(2007)28

OECD (2011) The OECD QSAR toolbox for grouping chemicals into categories, ECHA-11-L-08-EN, https://wwwqsartoolboxorg/documents/21638082/21638129/QSAR_Toolbox_leaflet_web_enpdf/405359d5-a844-4420-baae-ae102de5e92e. Accessed 2018–01-08

Palm Cousins A, Brorström Lundén E, Lexén J, Rydberg T (2018) Emissions from Articles. Synthesis report of the ChEmiTecs Research Program. Swedish Environmental Protection Agency REPORT 6802

Roos S, Holmquist H, Jönsson C, Arvidsson R (2017) USEtox characterisation factors for textile chemicals based on a transparent data source selection strategy. Int J Life Cycle Assess. https://doi.org/10.1007/s11367-017-1330-y

Rosenbaum RK, Bachmann TM, Gold LS, Huijbregts MAJ, Jolliet O, Juraske R, Koehler A, Larsen HF, MacLeod M, Margni M, McKone TE, Payet J, Schuhmacher M, van de Meent D, Hauschild MZ (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 13:532–546

Rybacka A, Andersson PL (2016) Considering ionic state in modeling sorption of pharmaceuticals to sewage sludge. Chemosphere 165:284–293

Rydberg T, Lexén J (2016) Comparing models in estimating additive emissions from plastic materials in use, platform presentation & extended abstract at SETAC Europe 26th Annual Meeting, Nantes 22–26 May, 2016

Saouter E, Aschberger K, Fantke P, Hauschild MZ, Bopp SK, Kienzler A, Paini A, Pant R, Secchi M, Sala S (2017a) Improving substance information in Usetox®, part 1: discussion on data and approaches for estimating freshwater ecotoxicity effect factors. Environ Toxicol Chem 36:3450–3462

Saouter E, Aschberger K, Fantke P, Hauschild MZ, Kienzler A, Paini A, Pant R, Radovnikovic A, Secchi M, Sala S (2017b) Improving substance information in usetox®, part 2: data for estimating fate and ecosystem exposure factors. Environ Toxicol Chem 36:3463–3470

TALETE (2014) TALETE srl, Dragon (software for molecular descriptor calculation) Version 6.0. http://www.talete.mi.it/. Accessed 2017–09-25

UNEP (2012) Global chemicals outlook towards sound management of chemicals trends and changes. Synthesis report for decision-makers. ISBN: 978-92-807-3275-7

US EPA (2012) EPI Suite™—estimation program interface. https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface. Accessed 2017–09-25

US EPA (2016) Toxicity Estimation Software Tool (TEST). https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test. Accessed 2017–09-25

van Leeuwen CJ, Vermeire TG (2007) Risk assessment of chemicals: an introduction. Springer, Netherlands. https://doi.org/10.1007/978-1-4020-6102-8

von der Ohe PC, Dulio V, Slobodnik J, de Deckere E, Kühne R, Ebert RU, Ginebreda A, de Cooman W, Schüürmann G, Brack W (2011) A new risk assessment approach for the prioritization of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the European Water Framework Directive. Sci Total Environ 409:2064–2077

Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, Öberg T, Dao P, Cherkasov A, Tetko IV (2008) Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. J Chem Inf Model 48:766–784