Accepted Preprint

Environmental Toxicology and Chemistry

Original Article

SPECIES SENSITIVITY DISTRIBUTIONS FOR USE IN ENVIRONMENTAL PROTECTION, ASSESSMENT AND MANAGEMENT OF AQUATIC ECOSYSTEMS FOR 12,386 CHEMICALS

LEO POSTHUMA, JOS VAN GILS, MICHIEL C. ZIJP, DIK VAN DE MEENT, DICK DE ZWART

Environ Toxicol Chem., **Accepted Article** • DOI: 10.1002/etc.4373

Accepted Article

"Accepted Articles" are peer-reviewed, accepted manuscripts that have not been edited, formatted, or in any way altered by the authors since acceptance. They are citable by the Digital Object Identifier (DOI). After the manuscript is edited and formatted, it will be removed from the "Accepted Articles" Web site and published as an Early View article. Note that editing may introduce changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. SETAC cannot be held responsible for errors or consequences arising from the use of information contained in these manuscripts.

This article is protected by copyright. All rights reserved

Accepted Preprint

Original Article

Environmental Toxicology and Chemistry
DOI 10.1002/etc.4373

SPECIES SENSITIVITY DISTRIBUTIONS FOR USE IN ENVIRONMENTAL PROTECTION, ASSESSMENT AND MANAGEMENT OF AQUATIC ECOSYSTEMS FOR 12,386 CHEMICALS

Running head: Species Sensitivity Distributions for 12,386 compounds

Leo Posthuma [a, b, *]

Jos van Gils [c]

Michiel C. Zijp [a]

Dik van de Meent [a,d,e]

Dick de Zwart [d,e]

This article contains online-only Supplemental Data

This article is protected by copyright. All rights reserved

Submitted 30 September 2018; Returned for Revisions 15 January 2019; Accepted 21 January 2019

This article is protected by copyright. All rights reserved

Accepted Preprint

Supplemental Data

Supplemental Data are available on the Wiley Online Library at DOI: 10.1002/etc.xxxx

Word count: 7823 words, excluding the word count in 40 references (1080 words) and including 5 Tables and 6 Figures and their respective captions.

Addresses and affiliations

[a] National Institute for Public Health and the Environment (RIVM), Centre for Sustainability, Environment and Health, P.O. Box 1, 3720 BA Bilthoven, the Netherlands. E: Leo.Posthuma@rivm.nl, T:+31-30-2742295;

[b] Radboud University, Department of Environmental Science, Institute for Water and Wetland Research, Faculty of Science, Radboud University, Nijmegen, The Netherlands;

[c] Deltares, P.O. Box 177, Delft, The Netherlands

[d] Mermayde, Groet, the Netherlands

[e] ARES, Odijk, the Netherlands

[*] Corresponding author. To whom correspondence may be addressed.

Word count: 7823 words, excluding the word count in 40 references (1080 words) and including 5 Tables and 6 Figures and their respective captions.

Author contribution: All authors contributed equally to this work

This article is protected by copyright. All rights reserved

Accepted Preprint

Abstract:

The present paper considers the collection and use of ecotoxicity data for risk assessment with Species Sensitivity Distributions (SSDs) of chemical pollution in surface water. SSDs are used to quantify the likelihood that critical effect levels are exceeded. This fits to the European Water Framework Directive, which suggest using models to assess the likelihood that chemicals affect water quality for management prioritization. We derived SSDs based on chronic and acute ecotoxicity test data for 12,386 compounds. The log-normal SSDs are characterized by the median and the standard deviation of log-transformed ecotoxicity data and by a quality score. A case study illustrates the utility of SSDs for water quality assessment and management prioritization. We quantified the chronic and acute mixture toxic pressure of mixture exposures for >22,000 water bodies in Europe for 1,760 chemicals for which we had both exposure and hazard data. Results show the likelihood of mixture exposures exceeding a negligible effect level and increasing species loss, respectively. The SSDs presented in this paper represent a versatile and comprehensive approach to prevent, assess and manage chemical pollution problems. This article is protected by copyright. All rights reserved

Key words: Aquatic ecotoxicity data, Chronic NOEC, Acute EC50, Species Sensitivity Distribution, Environmental risk assessment, Life cycle assessment

This article is protected by copyright. All rights reserved

Accepted Preprint

INTRODUCTION

Human activities cause the emissions of more than 100,000 chemical substances, with expected increases in compound diversity and emitted masses (Bernhardt et al. 2017; ECHA 2016; UNEP 2013). This results in diverse ambient concentrations (e.g., EMPODAT, http://www.normandata.eu/empodat/), body residues (USEPA 2009), ecological risks (Malaj et al. 2014) and eventual ecological and human health impacts (e.g., Hoekstra and Wiedmann 2014; Schäfer et al. 2016; Vörösmarty et al. 2010). Chemical pollution is a main driver of deterioration of freshwater biodiversity (Vörösmarty et al. 2010). Complementary policy approaches (chemical safety assessment, environmental quality assessment and management, and product environmental footprints) are used to prevent and limit impacts of such pollution. These require ecotoxicity data and a method to convert these data into estimates of benchmark concentrations for no- or negligible impacts (further on referred to as sufficiently protected) or in expected impact magnitudes of pollution (expressed as, e.g., species loss). Species Sensitivity Distributions (SSD) support making both these conversions, for separate compounds and mixtures (De Zwart and Posthuma 2005).

An SSD reflects the observation that inter-species differences in sensitivity to a chemical resemble a bell-shaped distribution (on a log-scale). An SSD is derived by fitting a selected statistical model (e.g., log-normal) to compound-specific ecotoxicity data. Lack of ecotoxicity data is often mentioned as reason that we have SSDs for only few chemicals for current policy applications. Criteria for SSD-data selection thereby vary amongst the policy approaches and jurisdictions, e.g. for the minimum number of data points (taxa) needed and for minimum study

This article is protected by copyright. All rights reserved

Accepted Preprint

quality characteristics (Posthuma et al. 2002). Despite that, SSDs are widely used for decision support (SI-Section 1). This likely relates to an observed association between SSD-predicted and observed biodiversity impacts (Posthuma and De Zwart 2012), to relative ease of use, and to representing a higher-tier approach as compared to using benchmark concentrations.

Several hundreds of thousands of ecotoxicity test results are available globally, but currently few are used to derive SSDs. Often strict criteria for SSD data selection (Klimisch et al. 1997; Moermond et al. 2016) and a minimum diversity of taxonomic groups and species (ECHA 2008) is prescribed in SSD derivation. That has resulted in currently low number of compounds with sufficient data to derive an SSD as well as to SSDs that are 'unstable' due to low data numbers.

In order to enable decision-support applications of SSDs for as many compounds and uses as possible, by reconsidering the aforementioned criteria, we collated species sensitivity data from existing sources. We created a base set of ecotoxicity data for test situations that may occur in nature. We were able to derive SSDs for a large number of compounds, and added a quality score to each SSD. We illustrate how SSDs and their quality scores are used in assessment planning and -interpretation.

The aims of the present paper are to describe:

1) data collection and curation for as many chemicals as feasible,
2) deriving chemical-specific SSDs, each with a quality score,
3) applying and evaluating methods to bridge data gaps,
4) utility and limitations of using SSDs in practical assessments (case study),

and

5) provide the set of SSDs for further use (SI-Table 2).

This article is protected by copyright. All rights reserved

Accepted Preprint

The scope for using the SSDs is global. SI-Table 2 presents two operationally defined SSD-models for the studied compounds, based on chronic no- or negligible effect data (e.g., NOEC, EC10, etc.) and acute median effective concentration (e.g, EC50) data, respectively. The former SSDs relate to current global practices in chemical safety assessment and environmental quality assessment and management, operating via protective benchmark concentrations (such as the Predicted No Effect (PNEC) and Environmental Quality Standards (EQS) (EC 2003; 2011), or similar benchmarks for geographies outside the European Union). Exposures below protective benchmarks are considered to imply no- or negligible impacts, and ecosystems are considered sufficiently protected at exposures below the benchmark. The latter SSDs relate to current global practices in Life Cycle Impact Assessment and other environmental assessments in which likely impacts of chemical pollution are quantified. That is commonly done in a comparative way, between products and environmental samples, respectively (see SI-Section 1). Increasing the number of compound-specific SSDs is relevant for all policy uses.

Abbreviations used in the text are summarized in Table 1.

MATERIALS & METHODS

Ecotoxicity data

Ecotoxicity data were collated from many sources, curated and operationally characterized for the two (chronic and acute) SSDs aimed at:

1. A validated set of existing and well referenced aquatic ecotoxicity database is described in De Zwart (2002). All available toxicity data were designated to represent chronic or acute toxicity criteria (Table 2):

This article is protected by copyright. All rights reserved

Accepted Preprint

a) records with the endpoints NOEC, LOEC, MATC, EC0, EC5, EC10 and EC20 are marked as "chronic NOEC" when they have an appropriate taxon-dependent test duration (see Table 2) and population relevant effect criterion (e.g., reproduction, growth, population growth, and development, next to mortality and immobility).

b) records with a sublethal (EC) or lethal (LC) endpoint ranging for 30% to 70% are marked as "acute EC50" when they have an appropriate taxon-dependent test duration (see Table 2) and effect criterion (e.g., mortality and immobility).

This dataset comprised of 30,806 records (3,445 substances; 1,556 different taxa; 2,513 chronic NOEC values; 28,293 acute EC50 values). As described by De Zwart (2002), this dataset was comparatively checked for plausible toxicity estimates. Implausible outcomes were traced to the original reference for data misinterpretations, and were often attributable to errors in unit transformations, typing errors and/or tests conducted under less optimal conditions. Erroneous entries were corrected when possible, and data were removed when original sources could not be checked (in this and later steps).

2. Further referenced data were added from an analysis and categorization of listed compounds of established and emerging concern under various national and international legislations (see https://www.stowa.nl/publicaties/ecologische-sleutelfactor-toxiciteit-hoofdrapport-deelrapporten-en-rekentools). Curation was as in dataset 1. Additionally, acute NOEC values and chronic EC50 values were identified using the test duration criteria.

   Data were obtained from a variety of sources:

   2.1. The USEPA ECOTOX database (https://cfpub.epa.gov/ecotox/data_download.cfm; download October 21st 2014). This addition comprised 58,714 records (1,853 substances;

This article is protected by copyright. All rights reserved

Accepted Preprint

942 taxa; 15,019 acute EC50 values; 19,875 acute NOEC values; 21,676 chronic NOEC values; 2,144 chronic EC50 values).

2.2. A total of 952 test results from fish embryo toxicity tests (FET) on 214 substances with 4 different fish species were adopted from the Procter & Gamble Company (Oris et al. 2012).

2.3. Das et al. (2013) and Sandersen et al.(2009) provided 334 records of measured acute EC50 toxicity concerning algae, daphnids and fish for 162 different pharmaceutical active ingredients.

2.4. A series of reports generated to define preliminary environmental quality criteria for compounds suspected to cause impact provided additional information mainly on pesticides and pharmaceuticals (Harezlak and Keijzers 2011; Osté et al. 2010; Smit and Keijzers 2015). This data addition consisted of 1,059 records on acute and chronic ecotoxicity for 37 substances involving 215 different taxa.

2.5. A series of Draft Assessment Reports (DAR) from EFSA (http://dar.efsa.europa.eu/dar-web/provision), and the Pesticide Properties Database (PPDB) (http://sitem.herts.ac.uk/aeru/ppdb/en/search.htm).

2.6. The Swiss Centre for Applied Ecotoxicology at EAWAG provided a large number of dossiers on the ecotoxicity of pharmaceuticals and pesticides (http://www.ecotoxcentre.ch/expert-service/quality-standards/proposals-for-acute-and-chronic-quality-standards/).

2.7. The WikiPharma database (http://www.wikipharma.org/welcome.asp, download November 18th 2016) was used to complement ecotoxicity data for listed pharmaceuticals.

This article is protected by copyright. All rights reserved

Accepted Preprint

3. Data in the REACH registry (April 24th 2015) after code harmonization added 207,943 records on acute and chronic toxicity on 8,787 different substances, mainly for algae, daphnids and fish. REACH data are not properly documented regarding test conditions, test duration and exposed taxa, and are not transparently traceable to peer reviewed origin. REACH-data were therefore analyzed both separately as well as after combining with the other data.
4. Read-across acute ecotoxicity data (baseline, acute LC50) for algae, daphnids and fish were derived for 5,201 substances. Toxicity estimates were evaluated as the lowest value derived by two different estimation methods: a) by ECOSAR prediction (Mayo-Bean et al. 2017), and b) methods utilized by UFZ, Leipzig, Germany. The UFZ method for acute fish toxicity consists of an automated read-across approach (Schüürmann et al. 2011). This model estimates baseline toxicity from log Kow and corrects it by a toxicity enhancement derived from experimental data for similar compounds. Compound similarity is deducted by comparison of atom-centered fragments (Kühne et al. 2009). For daphnids, a refined version of this approach has been applied (Kühne et al. 2013). For algae, a simple model similar to that for fish was used, but here the acute toxicity was directly derived from experimental data of similar chemicals taken from an internal data set. Again, these data were analyzed both separately as well as after combining with the other data.

SSD derivation

A variety of log-normal SSDs was derived. Available data were used to derive a compound-specific SSDs for all species tested. The location and scale parameters of log-normal SSDs (mu and sigma) were first derived with optimum data, and resulting SSDs were assigned a high

This article is protected by copyright. All rights reserved

Accepted Preprint

quality score (‘1111’). As decreasing test data numbers trades off into limited numbers of compounds and potentially lower SSD-‘stability’, additional SSDs were derived with various (inter-SSD) extrapolations to bridge data gaps, followed by an evaluation of similarity to the high quality ones.

In line with the data origin and the use contexts (both chemical safety- and environmental quality assessment), data were subdivided according to source quality (literature referenced versus non-referenced, and measured versus read-across) and to effect endpoints (chronic and acute) with characterization of data that were strictly measured ecotoxicity endpoints and obtained by extrapolation.

With the selection of strictly measured data, if more than two different taxa are tested for acute EC50 and chronic NOEC, the summarization process consisted of estimating the two moments of a log-normal SSD (Posthuma et al. 2002):

1) mu – the population median of toxicity values with equal weight per taxon, by first calculating the geometric average toxicity within taxa and subsequently calculating the geometric average toxicity over taxa of the geometric average toxicities per taxon, and
2) sigma – the population standard deviation of ^{10}log transformed toxicity data, without considering taxon weight.

If ecotoxicity data were available for less than 3 taxa, the process was restricted to mu - the average of the population of ^{10}log transformed toxicity values with equal weight per taxon. For the substances with too little data an intermediate median SSD slope (sigma) was adopted with a value of 0.7 (the average slope over all data is 0.71).

The above ‘strictly measured’ data selection leaves many of the collated data unused. When sufficient acute EC50 or chronic NOEC values are not available but other toxicity endpoints are,

This article is protected by copyright. All rights reserved

Accepted Preprint

those are extrapolated to acute EC50 and chronic NOEC values using empirical derived extrapolation values (De Zwart 2002; Duboudin et al. 2004). Mind that the correct interpretation of this extrapolation is a parallel shift in an SSD, which is far more robust than per-species Acute-Chronic Ratio extrapolations (De Zwart 2002). Further test endpoints on acute EC50 and chronic NOEC were firstly obtained after extrapolation from the most prominently available single other test endpoints in the collated data set. This was done similar to the procedure for the strictly reported ecotoxicity data. Extrapolation factors are summarized in Table 3. If more than 4 substances shared the same primary Mode of Action (MoA) in this extrapolation procedure, the SSD slope (sigma) was averaged over all substances with a similar primary MoA. If more than 5 different taxa are tested for either acute EC50 or chronic NOEC, the input data were not further extrapolated. If test data are available in this extrapolation process for less than 3 taxa, the test endpoints (acute EC50 and chronic NOEC) were summarized by extrapolation from all available data, irrespective of the reported test endpoint. This was done according to the scheme and the order presented in Table 3, as derived from De Zwart (2002), and approximately reconfirmed by Figure 3 in the present paper. If in the extrapolation process the maximum coverage of ecotoxicity data is available for less than 3 taxa, the summarization process is based on strictly measured data and restricted to mu - the average of the population of ^{10}log transformed toxicity values with equal weight per taxon.

SSD types and quality scores

The database contains far more acute EC50 data than chronic NOEC data, making SSDs from the former data more robust for the majority of compounds. There are also other reasons why some SSDs are likely better estimates of true but unknown assemblage-wide sensitivity differences

This article is protected by copyright. All rights reserved

Accepted Preprint

than others. Therefore, we added four-digit quality scores as shown in Table 4Table 4, with the modality '1111' representing the highest-quality. Derivation of the slope may or may not be possible (digit 1). Representation of different taxonomic groups was ranked (digit 2). Data origin was classified (digit 3). When SSDs were derived via data-bridging techniques, the method was scored (digit 4). The quality score information was added for planning and interpreting an assessment. This has a specific utility when large numbers of samples and compounds are assessed, and where an assessment commonly involves prioritization of management efforts to 'most impacted sites' and 'most contributing compounds'. That is, prioritization is straightforward if all SSDs used in an assessment are of high quality. If some SSDs are of low quality, this generally indicates a need for collecting additional hazard data. An exception may occur for assessments with limited resources, with clearly high-ranking cases and some low-ranking cases derived from (partially) lower-quality SSDs. Uncertainty analysis may reveal whether additional hazard data can 'move' cases with a low rank and a low SSD quality 'up' to the group prioritized for management attention.

Evaluation of SSDs

Various regressions of 'other' SSDs on high-quality ones were performed to evaluate quality of the 'other' SSDs (for mu, the SSD midpoint). Similar outcomes suggest that data sets can be merged to derive SSDs from more data per compound. The regressions involved REACH data, read-across data, and extrapolated acute EC50 or chronic NOEC data (cf. Table 3Table 3). These were regressed against the geometrical average of high-quality SSDs, defined as strictly measured acute EC50 or chronic NOEC toxicity data over tested taxa derived with the curated and validated data for the substances. Also, the geometric average of strictly measured chronic

This article is protected by copyright. All rights reserved

Accepted Preprint

NOEC toxicity data over tested taxa was regressed against the geometric average of strictly measured acute EC50 toxicity data.

Example case study: water quality assessment and management prioritizations

Scope. The case study was set up to examine the largest possible fraction of chemicals in commerce in Europe, focusing on water quality and pollution impacts.

Exposure assessment. Predicted Exposure Concentrations (PECs) were derived from EU-production data with an integrated EU-wide emission-fate-hydrological model (Van Gils et al. In preparation, 2019). Details on PEC derivation and -accuracy are in SI-Section 4. Combined with the SSD results, required exposure and hazard data were available for a subset of 1,760 compounds, selected for adequate physico-chemical and ecotoxicological data representation. PECs are freely dissolved concentrations and were derived for 22,278 modelled EU sub-catchments (median spatial resolution: 214 km^2) for a 365 day period (with weather data for the year 2013. This yielded $1.4 * 10^{10}$ PECs.

Impact assessments. Toxic pressures of the individual chemicals and their mixtures were derived from the PECs per sub-catchment/day using the SSDs of the present study. Compared to the mixture model of De Zwart and Posthuma (2005), we applied a simplified approach to derive mixture toxic pressures, by operationally assuming that all chemicals act concentration additively, in formula:

$$\left(\sum_{i=1}^{1760} HU_i, where\ HU_i\ is: \frac{Predicted\ environmental\ concentration_i}{SSD\ midpoint\ mu_i}\right).$$

This article is protected by copyright. All rights reserved

Accepted Preprint

with HU=Hazard Unit. This is similar to the frequent practice of water quality assessment for mixtures by linearly summed Risk Quotients over compounds, whilst numerical values estimated with the simplified and the original (more complex) model are similar. The approach was modeled with the MS-Excel function NORMDIST(^{10}log(ΣHU),0,Average SSD slope = 0.7,1). Chronic and acute mixture toxic pressures were quantified per sub-catchment/day.

Assessment and interpretation examples. The vast number of daily mixture toxic pressure data for individual subcatchments ($8.1 * 10^6$) were summarized using various approaches and statistics. Relative impact rankings across water bodies were visualized as GIS-maps, based on the years' P95 msPAF-values. Relative rankings of the contributions of chemicals to those impacts on a regional scale were derived in two steps. In the first step, a relative toxicity score for each compound in a sub-set of samples (Europe, or a specific example river basin) was determined as the product of the mean compound toxic pressure in the set and the ratio between the non-zero values for that compound and the non-zero values for the mixtures. Those represent the magnitude and the relative contribution and frequency of increased compound pressures to total mixture pressures. All scores were then relatively ranked using the lowest-ranking compound as the baseline (defined as '1'). Case study example data are shown for all European basins, or for some assessments for a typical northwestern (Rhine), a southeastern (Danube) and a set of southern basins (the Spanish basins of Ebro, Guadalquivir, Xuquer and Llobregat, combined), respectively.

This article is protected by copyright. All rights reserved

Accepted Preprint

RESULTS

Ecotoxicity data

The collated and curated data set for deriving SSDs consists of 256,409 records (details on the data set are in SI-Section 2). Data origins (strictly measured and referenced up to read-across) were tagged, to allow derivation and comparisons of various types of SSDs for a compound.

SSD types and quality scores

A single compound may have various SSDs: e.g., from chronic or acute data, from referenced data, unreferenced REACH data or read-across data, and from strictly measured or extrapolated data. SSDs could be derived for 12,386 compounds, where 12.214 originate from acute EC50 data and 7540 from chronic NOEC data. Their characteristics are summarized in SI-Section 3 (SI-Table 2), where the acute and chronic SSD data are selected to represent the lowest available quality score (best quality SSD), while combining the referenced and unreferenced REACH data. The quality scores vary from '1111' to '2436' or '2411' for acute and chronic SSDs, respectively. SI-Table 2 also contains a summary overview of the numbers of compounds per SSD-type.

Evaluation of SSDs

The first evaluation considered regressions of acute SSDs derived from various methods ('other') on high-quality non-extrapolated SSD-midpoints. The regressions were all significant (Figure 1**Error! Reference source not found.**): SSDs based on acute EC50 REACH data (A), overlapping number of compounds 927, $y = 0.8173x + 0.7025$, $R^2 = 0.65$, $P < 0.001$; SSD based on acute EC50 read-across data (B), overlapping number of compounds 1,116, $y = 0.5914x + 1.2515$, $R^2 = 0.49$, $P < 0.001$; SSD based on acute EC50 data extrapolated from other test endpoints (C), overlapping number of compounds 3,827, $y = 0.9611x + 0.1362$, $R^2 = 0.95$, $P <$

This article is protected by copyright. All rights reserved

Accepted Preprint

0.001). The read-across SSDs (related to baseline-toxicity) showed more variability, and appeared to underestimate the mu for the most toxic substances up to a factor of 10 as compared to high-quality SSDs. The reliability of acute-EC50 SSDs based on the other data analysis methods decreases in the sequence Extrapolated SSDs ≈ REACH-based SSDs > Read-across SSDs.

Insert Figure 1.

The second evaluation similarly considered two regressions, now for SSDs from chronic NOEC data (Figure 2**Error! Reference source not found.**). For chronic NOEC REACH-data, the relationship with the chronic high-quality non-extrapolated SSD-midpoint data was less strong than for the EC50-data, but still highly significant (A, overlapping number of compounds 251, $y = 0.7448x + 0.8502$, $R^2 = 0.60$, $P < 0.001$). For the most toxic substances, the REACH data tend to underestimate toxicity as compared to the well-referenced dataset by a factor of up to approximately 30. A highly significant relationship was found between extrapolated chronic data and the chronic high-quality non-extrapolated SSD-midpoint data (B, overlapping number of compounds 1,131, $y = 0.7941x+0.5902$, $R^2 = 0.74$, $P < 0.001$). The extrapolated data also tend to underestimate toxicity as compared to the high-quality data by a factor of up to approximately 10. The reliability of chronic-NOEC SSDs based on the other data analysis was relatively similar for both other methods.

Insert Figure 2.

The third evaluation considered the comparison of SSDs based on NOECs versus EC50s. This showed a significant association between the mu values for the 250 substances for which both aspects are strictly measured and quantified (Figure 3**Error! Reference source not found.**, n =

This article is protected by copyright. All rights reserved

Accepted Preprint

250, y = 0.7304x - 0.4414, $R^2 = 0.60$, $P < 0.001$). On average, the chronic NOEC is less than a factor of 10 lower than the acute EC50, which represents an observed factorial shift of the SSD. A limited number of outliers have a strong influence on this regression. A restriction of the regression to the 5-95 percentile data intervals yielded a factorial difference of about 6.6 and a slope of about 0.85 (n = 225, y = 0.8509x - 0.8237, $R^2 = 0.76$, $P < 0.001$). These outcomes suggest that, on average, an SSD-NOEC can be derived from an SSD-EC50 by extrapolation (as shown in Table 3Table 3), but also that the variability around the average should be taken into account when such an extrapolation would be used in practice.

Insert Figure 3

Example case study: water quality assessment and management prioritizations

Scope and illustrative purpose. Presented results illustrate how the method allows for impact rankings of sites and of relative importance of substances within basins or water bodies. The impact assessments are based on the large series of exposure concentrations (detailed in SI-Section-4). It should be noted that alternative data summary choices – related to the assessment problem – yield different results. Results are therefore explicitly not to be interpreted as a list of Europe-wide priority sites or priority chemicals in a WFD context.

Impact assessment and prioritizations. Various impact assessment and prioritization outcomes are illustrated, based on the large set of PECs for 1760 substances that all have relatively high SSD quality scores, from ‘1111’ to ‘1324’or ‘1325’ for chronic and acute, respectively. Mixture toxic pressures based on chronic or acute SSDs were derived and mapped, to illustrate the spatial variation of these impact-related metrics (Figure 4 and Figure 5, respectively). As GIS-maps cannot plot both space and time, the example Figures are based on the 95th percentiles of

This article is protected by copyright. All rights reserved

Accepted Preprint

predicted mixture toxic pressures over a year (P95-year = 18 days). The mixture toxic pressure of a water body is higher for 5% of days in the modelled year. The interpretation proceeds as follows:

(1) Per-species interpretation. Colors represent the variation of the probability that a randomly selected species from the set of tested species would be exposed beyond the species' chronic no-effect level or the acute EC50 for at most 18 days per year. These outcomes represent the 'Probability of Effects on a Species' (PES) of the water pollution at the levels of 'initiation' or 'substantial' harm', respectively (Suter et al. 2002). PES-values characterize the potential of the polluted water to cause harm, which is the basis for the term 'toxic pressure'.

(2) Biodiversity interpretation. A quantitatively identical expression of the outcome is the Potentially Affected Fraction of species (PAF), but the interpretation narrative differs. The PAF expression relates to the regulatory-defined endpoint of protecting against biodiversity reduction in species assemblages. PAF values relate to the concept of protective benchmarks, if those are derived from an SSD of chronic NOECs, as HC5 (Hazardous Concentration for 5% of the species). Higher toxic pressures imply higher fractions of species affected for the studied acute or chronic endpoint.

(3) Regulatory interpretation. The two maps relate to two current policy approaches, viz. chemical safety- and environmental quality assessment policies, and ecological impact assessment. The former two approaches operate via the protective benchmark no-effect concept (using PNEC and EQS for REACH and the WFD, respectively), below which ecosystems are considered 'sufficiently protected' for expected or observed exposures. For single compounds or mixtures sufficient protection relates to $PAF\text{-}NOEC_{max}$ = $msPAF\text{-}NOEC_{max}$ = 0.05, which is regulatory considered to protect 95% of the species against adverse effects. In Figure 4**Error!**

This article is protected by copyright. All rights reserved

Accepted Preprint

Reference source not found., this equivalency was used as class boundary between sufficient and insufficient protection (blue-green boundary), with other colors represent increasing distress (exposure higher than the no-effect level). In Figure 5**Error! Reference source not found.**, the color scale relates to increased biodiversity effects, found in empirical studies, which can be aligned with ecological impact classification used in the WFD to define moderate, poor and bad water quality.

Insert Figure 4

Insert Figure 5

Impact distributions across water bodies were investigated with the same data. The water quality would be classified as 'insufficient protected' for approx. 65% of all European water bodies (P95-year msPAF-NOEC data, Figure 6**Error! Reference source not found.**). Specific basins have higher fractions of water bodies with 'insufficient protection', with average values of ≈ 93%, 88% and 79% for the example basins of the Rhine, the Danube and the Spanish basins, respectively. The observed Pareto-type (skewed) distributions imply that relatively few sites are characterized by relatively high chronic toxic pressures.

Insert Figure 6

Relative impact contributions of chemicals were investigated, with the same data (example for acute-SSD EC50 ranking), and toxic pressure data aggregated over an area and over time. We derived a top-15 ranking of substances (Table 5

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 5). The top-15 explained nearly 99.5% of the mixture exposure effects, with <0.5% explained by the remaining 1,745 compounds. This Pareto-type distribution implies that approx. 1% of the compounds causes 99% of the exposure impacts, leading to a Pareto-type '99-1' rule for species loss for the P95-year assessment.

Some of the top-15 chemicals were not identified as potentially relevant following current compound prioritizations according to the WFD or the NORMAN network (kindly provided by Valeria Dulio, Executive Secretary of NORMAN; https://www.norman-network.com/sites/default/files/files/suspectListExchange/NORMAN_PriorityList_2016.csv). Modeling expected impacts may identify chemicals that likely affect ecosystems, but that are not identified by monitoring (due to lack of attention, lack of analysis methods, or detection limits that are higher than effect benchmarks).

Final interpretation of an assessment. In a comprehensive assessment, the above types of results would be further checked before results are used for management prioritization. First, ranking outcomes should be checked on their fit to the assessment problem. For example, prioritizations will differ when chemicals with peak exposures (such as pesticides) are involved, and the assessor may then evaluate outcomes from P99-year based ranking. The outcomes of this are illustrated in SI-Section 5. Second, SSD-quality scores should be evaluated for their potential influence on the interpretation (quality scores were high for the top-15 chemicals). Third, outcomes will normally be interpreted with other lines of evidence. In doing so, collection of data on the top-15 chemicals showed that two compounds, Terbufos and Phorate, are not any more approved in the EU. As all our assessment runs were based on dossier data which we did not *a*

This article is protected by copyright. All rights reserved

Accepted Preprint

priori screened on forbidden compounds, so that the high ranks are not relevant for *current* prioritizations; however, the *past* non-approval decision is supported by the high ranks found for these compounds. Further compound details are in SI-Section 6, showing that all top-15 compounds are characterized by high production mass, ubiquitous use, and high hazard classifications.

DISCUSSION

Key innovations

This study addresses the problem that different environmental policy frameworks have developed very different practices in handling ecotoxicity data and assessment models, with major trade-offs for practical assessments. Various guidance documents prescribe strict criteria for data selection and SSD derivation. The downside of such criteria is that the risks of most compounds and their mixtures cannot be evaluated, prioritized and managed. Current water quality assessment under the WFD considers e.g. only 0.2% of the compounds in commerce (EQSs for approx. 300 compounds (EEA 2018) versus 146,000 registered compounds on the REACH website). To address that trade off, we developed SSDs and associated mixture approaches for a large number of chemicals, to enable more comprehensive and realistic assessments.

Innovative aspects of the present study are (1) the methods and sources for the collection and curation of ecotoxicity data, (2) the SSD-derivation and quality scoring method, (3) the extra information that can be gained from inter-SSD comparisons and extrapolations, (4) the utility of SSD-based assessment outcomes for ranking sites and compounds (illustrated in the case study), and eventually (5) the opportunity to use a consistent set of ecotoxicity data and SSDs for various practical policies. For the practical uses, assessors should be aware of the limitations of SSD-

This article is protected by copyright. All rights reserved

Accepted Preprint

based assessments, as SSDs do not consider food-chain exposures or indirect effects (e.g, on predators via toxicity to prey). Moreover, they should be aware of the fact that the *simple* expression of the toxic pressure for a water body has the *complex* interpretation that the species that are exposed will show widely different impacts, related to the species sensitivity differences that are the basis of SSDs.

Expanded number of compounds, SSDs, quality scores and SSD applications

We operationally derived separate acute and chronic SSDs for many compounds. Those may – in principle – be used for all policy purposes. Checking SSD-quality scores is thereby always important. Low quality scores (e.g, caused by few ecotoxicity data or by extrapolation) may be consequential. For example, the calculated acute-median sensitivity (mu-acute) may be smaller than the calculated chronic value (mu-chronic) due to the haphazard effect of small data sets. This specific effect occurred for ≈ 1.3% of the SSDs (160 data lines in SI-Table 2). Low SSD quality scores can only be improved by collecting more test data. When quality scores are considered sufficient, the way of using the SSDs for practical assessments is basically as follows. First, we acknowledge that current guidances exist on deriving protective benchmark concentrations, and results of the present study may therefore not be adopted for specific policies. However, if there are no data that fit the current guidance for a contaminant of potential concern, a provisional benchmark concentrations can be derived, e.g., via the chronic-acute SSD-level relationship shown in Figure 3, to provide provisional insights in potential impacts for data-poor chemicals. Second, the utility for water quality assessment and management prioritization was shown in the case study, based on relative rankings of impact levels across water bodies and compounds within the studied basins. Third, the SSDs can be used for establishing the

This article is protected by copyright. All rights reserved

Accepted Preprint

environmental (ecotoxicity) footprints of products as well as production-chain ecotoxicity hot spots, to enable production and selection of environmentally benign products. The European Union recently derived SSD-based effect factors for this purpose from three regulatory data sources for thousands of chemicals (Saouter et al. 2018). Consistent environmental protection and pollution management can be based on the SSDs presented in this study, for a wide array of compounds. Critical use must be supported by the quality scores. If needed, specific SSD types with specific quality scores may be earmarked for specific purposes by the user, e.g. for repetitive assessments.

Prioritization opportunities

All assessment outcomes of SSDs relate to ranking, as basis for management prioritization and -efficacy. For an array of compounds, the rank order of protective benchmarks (e.g., HC5s, PNECs, EQSs) reflects the relative potency of different compounds to cause harm. For an array of sites, the order of (mixture) toxic pressures similarly reflects impact differences across sites. For an array of products, outcomes identify benign products and production chain hotspots. Outcomes of acute-data SSDs have even been used in disaster assessment and management by UNDAC teams (SI-Section 1). The case study only illustrates the ranking of polluted water bodies. Based on experiences with case study data (such as in Table 5

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 5), we recommend that rankings for all applications are not interpreted as absolute and as fixed order of cases, as natural and man-made variability of exposures occur. That is, pesticides may have a high rank-order only during the growing season, with further influences of weather (e.g. rainfall events) affecting emissions of some chemicals (via run-off) and dilution of all chemicals. Instead of assuming an absolute idea of site- or substance ranking, we propose to use classes. For example, 1) an 'always high'-ranking class of sites or compounds where the probability of impact is always high, 2) an 'always low'-ranking class, where the low probability of impact allows to 'exclude the innocent' and 3) an 'intermediate' class where the probability of impact depends on the situation. These classes would discriminate three clear management perspectives, 1) action needed, 2) no action needed, and 3) possible further lines of evidence needed. The concept of using classes is further supported by results from field monitoring (Vallotton and Price 2016). Based on this phenomenon, assessors can also consider the opposite of prioritizing the higher-impact sites or compounds for management, by considering the lower tail of the distribution (Figure 6**Error! Reference source not found.**). With a large number of ranked cases and limited management resources, there may be an opportunity to 'neglect the innocent', even when some SSD-quality scores underlying the lower-tail ranking are low.

Case study: utility example on ranking toxic pressures for sites and substances

The case study illustrates how site and substance ranking provides management prioritization insights, refining global water quality classification practices. The latter are currently based on comparing single-chemical exposure concentrations to a protective benchmark. This results in a binary outcome of water quality assessment: there is '(in)sufficient protection', which is then interpreted and communicated as 'polluted' or 'unpolluted'. When applied to mixtures, the application of such an approach to European surface waters showed that all water bodies of a

This article is protected by copyright. All rights reserved

Accepted Preprint

country can be interpreted as polluted (i.e., Sweden, EEA 2012). This incorrectly seems to suggest equal management needs for all studied water bodies, neglecting that low and high benchmark exceedance implies lower and higher impacts, and lower or higher motives for pollution reduction. The case study shows that current practices can be refined, to highlight where impacts are likely highest, due to which compound(s)/groups. These rankings help prioritization of management, and select measures that reduces impacts (improves water quality) most. Novel case study insights are as follows. First, European water quality is currently insufficiently protected (Figure 4 Error! Reference source not found.**), corroborating the study of Malaj et al. (2014). Second, this implies an associated degree of likely species loss (Figure 5**Error! Reference source not found.**), based on empirical evidence for the association between msPAF-acute EC50 and species loss (e.g., Posthuma et al. 2016). Third, this is attributable to relatively few compounds at the European- and basin scales (Table 5**

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 5), also found by others (e.g., Vallotton and Price 2016). Fourth, the high-ranking compounds share specific characteristics: high mass used in Europe and ubiquitous use and high-hazard characteristics for aquatic ecosystems (see SI-Section 6). This means that chemical safety assessment knowledge (as collected for e.g. REACH) provides meaningful indications of potential impacts in aquatic ecosystems. All these insights were obtained with compounds with high SSD-quality scores. The sequence of analysis steps was designed in line with the holistic principles of the WFD. It implies a stepwise and meaningful "System-Site-Substance-Solution" focus in the assessment of impacts and planning of management of European surface waters.

Assessment problem definition, model choice and interpretation

The case study resulted in outcomes for a specific set of conditions (e.g, year P95 data). In practical assessments, the most-informative outcomes for management prioritization should be generated by tailoring the assessment to the specific conditions. For example, if the emissions of all chemicals in a region are rather constant (e.g. household chemicals), the assessor may investigate especially spatial exposure and impact ranking using e.g. the year-P50 or P95 toxic pressures, whereas for exposures in an agricultural landscape the assessor may focus on peak exposures (e.g. year P99 data). SI-Section 5 illustrates the increased toxic pressures when using e.g. the P99-year data (4 days peak exposures are covered). The number of sufficiently protected sites (chronic mixture toxic pressure < 0.05) is lower, and predicted species loss is higher. It is important to note that the applied EU-wide model operates with a spatial resolution of approximately 200 km^2, and consequently, probabilities of impacts at point sources (e.g., downstream waste water treatment plants) are not shown in detail. In the vicinity of point

This article is protected by copyright. All rights reserved

Accepted Preprint

sources, impacts may be much higher than shown in this study via water-body level exposure data.

Combining lines of evidence and planning of monitoring

An assessment of the likelihood of impacts under the WFD consists of various lines of evidence. Model results can be combined with information from an 'assessment of pressures' (human activities), available monitoring data, and other information. Information may consider sources as diverse as non-target screening of the presence of chemicals (Hollender et al. 2017) to effect-based methods that assess impacts of complex mixtures in water samples (Altenburger et al. 2015). Planning and management of river basins combines these lines of evidence, and the use of SSDs in this process allows the assessor to obtain highly specific information on the likelihood that chemical pollution causes harm.

The case study results also contain a suggestion on monitoring and management for ten additional compounds as compared to current practices (Table 5

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 5).

Communicating results and evaluating trends

Currently, one of the conundrums of pollution assessment and management is the communication of results. In the case study, $1.5 * 10^{10}$ exposure concentrations had to be summarized for management planning, and this number multiplies if an assessor wants to evaluate trends of past management or of optional abatement strategies. To address this problem, mixture toxic pressures can be summarized as chemical footprints for a region (Bjørn et al. 2014; Zijp et al. 2014). The changes caused by past management or future abatement scenarios can then be summarized and communicated via chemical footprints, to summarize spatial or temporal trends in mixture impacts for large regions. Thus, SSDs can be used as an effective, though lower-tier (screening) approach for water quality assessment and management in the context of a wide diversity of policy fields, with opportunities to explore 'big patterns' (footprints) as well as 'details' (specific sites and chemicals within sites). The use of SSDs provides an intermediary tool that lies between generic assessments of chemical safety and more specific impacts assessments based on more complex (ecological) modeling.

CONCLUSIONS

1. Species Sensitivity Distributions (SSD) are used in environmental protection, assessment and management practices, currently for a few to a few thousand compounds only
2. SSDs are provided for 12,386 substances, with a quality score to assist in planning and interpretation of assessments
3. The utility of the SSDs was illustrated for water quality assessment at the European scale considering 1,760 compounds and their mixtures

This article is protected by copyright. All rights reserved

Accepted Preprint

4. The role of chemical pollution on aquatic ecosystems could be specified, regarding both the regulatory concept of sufficient protection (REACH and WFD) as well as species loss (WFD, ecological status impact classification)
5. The use of models is suggested in the European WFD, and SSDs are fit for that use as they help to express expected impact magnitudes of pollution
6. The use of SSDs for water quality assessment follows the holistic principles of the WFD, and supports a "System-Site-Substance-Solution" sequence in the assessment of impacts and planning of management
7. The use of SSDs complements the current per-chemical benchmark approach, which substantially improves the diagnosis and communication of chemical pollution in surface waters

SUPPLEMENTAL DATA

Supplemental Data are available on the Wiley Online Library at DOI:. 10.1002/etc.xxxx

ACKNOWLEDGEMENT

This study was funded by SOLUTIONS and co-funded by the Dutch Foundation for Applied Water Research (STOWA) in the context of project "Development of the Ecological Key Factor Toxicity" [Ecologische Sleutel Factor Toxiciteit]. The SOLUTIONS project is supported by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement no. 603437. Data source REACH data: European Chemicals Agency, http://echa.europa.eu/. The data were extracted from IUCLID on April 24th 2015 as granted by the competent authorities, and data treatments have been done respecting pertinent
This article is protected by copyright. All rights reserved

Accepted Preprint

confidentiality rules. Only trained and qualified personnel was authorized to work on the downloaded data, to derive SSDs. Ralph Kühne is gratefully acknowledged for expanding the toxicity data base. E-HYPE generated hydrology data used for the case study were provided to the SOLUTIONS project by SMHI, Sweden. The valuable remarks of peer reviewers, ET&C editor Mirco Bundschuh, Kees Kramer and Jappe Beekman to earlier drafts of this manuscript are greatly acknowledged.

REFERENCE LIST

Altenburger R, Ait-Aissa S, Antczak P, Backhaus T, Barceló D, Seiler TB, Brion F, Busch W, Chipman K, de Alda ML et al. 2015. Future water quality monitoring - adapting tools to deal with mixtures of pollutants in water resource management. Science of the Total Environment. 512-513:540-551.

Bernhardt ES, Rosi EJ, Gessner MO. 2017. Synthetic chemicals as agents of global change. Frontiers in Ecology and the Environment. 15(2):84-90.

Bjørn A, Diamond M, Birkved M, Hauschild MZ. 2014. Chemical footprint method for improved communication of freshwater ecotoxicity impacts in the context of ecological limits. Environmental Science & Technology. 48(22):13253-13262.

Das RN, Sanderson H, Mwambo AE, Roy K. 2013. Preliminary studies on model development for rodent toxicity and its interspecies correlation with aquatic toxicities of pharmaceuticals. Bulletin of Environmental Contamination and Toxicology. 90(3):375-381.

De Zwart D. 2002. Observed regularities in species sensitivity distributions. In: Posthuma L, Suter GW, Traas TP, editors. Species sensitivity distributions in ecotoxicology. Boca Raton: Lewis. p. 133-154.

De Zwart D, Posthuma L. 2005. Complex mixture toxicity for single and multiple species: Proposed methodologies. Environmental Toxicology and Chemistry. 24(10):2665-2676.

Duboudin C, Ciffroy P, Magaud H. 2004. Acute- to- chronic species sensitivity distribution extrapolation. Environmental Toxicology and Chemistry. 23(7):1774-1785.

EC. 2003. Technical guidance documents on risk assessment, part ii. Eur 20418 en/2 (https://echa.Europa.Eu/documents/10162/16960216/tgdpart2_2ed_en.Pdf). Ispra, Italy: European Chemicals Bureau, Joint Research Centre.

EC. 2011. Common implementation strategy for the water framework directive (2000/60/ec) - guidance document no. 27 - technical guidance for deriving environmental qualilty standards. Brussel, Belgium.

ECHA. 2008. Guidance on information requirements and chemical safety assessment. Helsinki, Finland: European Chemicals Agency.

This article is protected by copyright. All rights reserved

Accepted Preprint

ECHA. 2016. Information on chemicals. [cited 2018, march 6]. Available from: Http://echa.Europa.Eu/information-on-chemicals. Europan Chemicals Agency. Helsinki, Finland.

EEA. 2012. European waters - assessment of status and pressures. Copenhagen, Denmark: European Environment Agency. No. 8/2012.

EEA. 2018. European waters - assessment of status and pressures 2018. Copenhagen, Denmark: European Environment Agency. No. 7/2018.

Harezlak V, Keijzers R. 2011. Afleiding van het ecotoxicologisch deel voor 29 ad hoc mtr's voor 2010. Delft, The Netherlands: Deltares. No. 1203121-001.

Hoekstra AY, Wiedmann TO. 2014. Humanity's unsustainable environmental footprint. Science. 344(6188):1114-1117.

Hollender J, Schymanski EL, Singer HP, Ferguson PL. 2017. Nontarget screening with high resolution mass spectrometry in the environment: Ready to go? Environ Sci Technol. 51(20):11505–11512.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology. 25(1):1-5.

Kühne R, Ebert R-U, Schüürmann G. 2009. Chemical domain of qsar models from atom-centered fragments. Journal of Chemical Information and Modeling. 49(12):2660-2669.

Kühne R, Ebert R-U, von der Ohe PC, Ulrich N, Brack W, Schüürmann G. 2013. Read-across prediction of the acute toxicity of organic compounds toward the water flea daphnia magna. Molecular Informatics. 32(1):108-120.

Malaj E, von der Ohe PC, Grote M, Kühne R, Mondy CP, Usseglio-Polatera P, Brack W, Schäfer RB. 2014. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proceedings of the National Academy of Sciences. 111(26):9549–9554.

Mayo-Bean K, Moran-Bruce K, Meylan W, Ranslow P, Lock M, Nabholz V, Von Runnen J, Cassidy L, Tunkel J. 2017. Methodology document for the ecological structure-activity relationship model (ecosar) class program v2. Washington DC, USA: Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency.

Moermond CTA, Kase R, Korkaric M, Ågerstrand M. 2016. Cred: Criteria for reporting and evaluating ecotoxicity data. Environmental Toxicology and Chemistry. 35(5):1297-1309.

Oris JT, Belanger SE, Bailer J. 2012. Baseline characteristica and statistical implications for the oecd 210 fish early-life stage chronic toxicity test. Environmental Toxicology and Chemistry. 31(2):370–376.

Osté L, Keijzers R, Mastalerz V. 2010. Afleiding van het ecotoxicologisch deel voor 28 ad hoc mtr's voor 2009. Delft, The Netherlands: Deltares. No. 1203121-001.

Posthuma L, De Zwart D. 2012. Predicted mixture toxic pressure relates to observed fraction of benthic macrofauna species impacted by contaminant mixtures. Environmental Toxicology and Chemistry. 31(9):2175–2188.

Posthuma L, De Zwart D, Keijzers R, Postma J. 2016. Ecologische sleutelfactor toxiciteit. Deel 2: Kalibratie: Toxische druk en ecologische effecten op macrofauna. Amersfoort, the Netherlands: STOWA.

Posthuma L, Suter GWI, Traas TP. 2002. Species sensitivity distributions in ecotoxicology. Boca Raton, FL, U.S.A.: CRC-Press.

This article is protected by copyright. All rights reserved

Accepted Preprint

Sanderson H, Thomsen M. 2009. Comparative analysis of pharmaceuticals versus industrial chemicals acute aquatic toxicity classification according to the united nations classification system for chemicals. Assessment of the (q)sar predictability of pharmaceuticals acute aquatic toxicity and their predominant acute toxic mode-of-action. Toxicology Letters. 187(2):84-93.

Saouter E, Biganzoli F, Ceriani L, Versteeg D, Crenna E, Zampori L, Sala S, Pant R. 2018. Environmental footprint: Update of life cycle impact assessment methods – ecotoxicity freshwater, human toxicity cancer, and non-cancer. Luxembourg: European Union. ISBN 978-92-79-98182-1, doi: 10.2760/178544 No. JRC114227.

Schäfer RB, Kühn B, Malaj E, König A, Gergs R. 2016. Contribution of organic toxicants to multiple stress in river ecosystems. Freshwater Biology. 61(12):2116–2128

Schüürmann G, Ebert R-U, Kühne R. 2011. Quantitative read-across for predicting the acute fish toxicity of organic compounds. Environmental Science & Technology. 45(10):4616-4622.

Smit CE, Keijzers R. 2015. Indicatieve waterkwaliteitsnormen voor bestrijdingsmiddelen - normvoorstellen voor 19 stoffen. Bilthoven, the Netherlands: RIVM. No. Briefrapport 2015-0124.

Suter GW, Traas TP, Posthuma L. 2002. Issues and practices in the derivation and use of species sensitivity distributions. In: Posthuma L, Suter GWI, Traas TP, editors. Species sensitivity distributions in ecotoxicology. 1st ed. Boca Raton, FL, U.S.A.: CRC Press. p. 437-474.

UNEP. 2013. Global chemicals outlook. Towards sound management of chemicals. Geneva, Switzerland: United Nations Environment Programme.

USEPA. 2009. The national study of chemical residues in lake fish tissue. Washington, DC: U.S. Environmental Protection Agency, Office of Water.

Vallotton N, Price PS. 2016. Use of the maximum cumulative ratio as an approach for prioritizing aquatic coexposure to plant protection products: A case study of a large surface water monitoring database. Environmental Science and Technology. 50(10):5286-5293.

Van Gils J, Schueder R, Chrzanowski C, Markus A, Cousins I, Lindim C, Focks A, Baveco H, Posthuma L, Kutsarova S et al. In preparation, 2019. An integrated model system for europe-wide assessment of the risks to ecosystems and human health of emerging pollutants. Environmental Toxicology and Chemistry.

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, Glidden S, Bunn SE, Sullivan CA, Liermann CR et al. 2010. Global threats to human water security and river biodiversity. Nature. 467(7315):555-561.

Zijp MC, Posthuma L, Van de Meent D. 2014. Definition and applications of a versatile chemical pollution footprint methodology. Environmental Science and Technology. 48(18):10588−10597.

This article is protected by copyright. All rights reserved

Accepted Preprint

FIGURE CAPTIONS

Graphical Abstract. Graph illustrating our contribution to an increased ability to provide environmental decision support with Species Sensitivity Distributions.

Figure 1. Comparison of acute-EC50 SSD midpoint data for three SSD-derivation approaches regressed on highest-quality SSD-midpoint data. X-data: highest-quality SSDs (Traceable data, All taxa, Geometrically averaged, Measured data, Acute EC50). Y-data types: 'other' and 'derived' SSDs (A=Untraceable data [REACH], blue; B=Read-across, green; C=Optimally extrapolated, red).

Figure 2. Comparison of chronic NOEC SSDs midpoint data for two SSD-derivation approaches regressed on highest-quality midpoint data. X-data: highest-quality SSDs (Traceable data, All taxa, Geometrically averaged, Measured data, Chronic NOEC). Y-data types: other SSDs (A=Untraceable data [REACH], B= Optimally extrapolated, red).

Figure 3. Comparison of SSD-midpoints for acute EC50 (X) and chronic NOEC (Y). A: All data points, blue dots and blue line. B: Ibidem, restricted to the P5-P95 data ranges, red dots and red line.

This article is protected by copyright. All rights reserved

Accepted Preprint

Figure 4. Results mapped in relation to the regulatory concept of 'sufficient protection' of aquatic ecosystems (initial effects, distress), used in chemical safety assessment (e.g, REACH: PNEC) and water quality assessment of chemicals according to the WFD (EQS). Blue: species assemblages in the water bodies are 'sufficiently protected' for 95% of the days (msPAF-NOEC <0.05, P95 of a year=18 days (see text). Technical output specification: numbers are P95-year of msPAF-chronic NOEC per site (n = 22,728 water bodies).

Figure 5. Results mapped in relation to the regulatory concept of ecological impact magnitudes (species loss), as utilized to classify ecological impacts under the WFD. Higher msPAF-acute EC50 values empirically relate to increased species loss. Technical output specification: numbers are P95-year msPAF-acute EC50 per site (n = 22,728 water bodies).

Figure 6. Cumulative distributions of mixture toxic pressures of the individual sub-catchments of Europe (n = 22,728) and of three example basin areas (Rhine: n = 813; Danube n = 3477; Spanish basins n = 696). X-axis = rank-order of the sub-catchments. Y-axis: probability that exposures exceed the no-effect- or 50%-effect level, under the assumption of concentration additivity of the substances in the mixtures.

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 1. Glossary of acronyms and abbreviations used.

Abbreviation	*Meaning*
ACR	Acute/Chronic Ratio
DAR	Draft Assessment Reports from EFSA
EAWAG	Eidgenössische Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz (Swiss Federal Institute for Environmental Science and Technology)
EC0	Zero percent Effect Concentration
EC10	Ten percent Effect Concentration
EC20	Twenty percent Effect Concentration
EC5	Five percent Effect Concentration
EC50	Fifty percent Effect Concentration or Median Effective Concentration
ECOSAR	Predictive model based on ECOlogical Structure Activity Relationships
EFSA	European Food Safety Authority
E-HYPE	European Hydrological Predictions for the Environment modeled by the Swedish Meteorological and Hydrological Institute (SMHI)
EQS	Environmental Quality Standard (protective benchmark used in the EU-WFD)
EU	European Union
FET	Fish Embryo Toxicity test
HC5	Hazardous Concentration for 5 percent of species
HU	Hazard Units (Concentration divided by SSD midpoint concentration)
LOEC	Lowest Observed Effect Concentration
MATC	Maximum Acceptable Toxicant Concentration
MoA	Mode of Action
msPAF	Multi-substance Potentially Affected Fraction of species for mixtures of chemicals
NOEC	No Observed Effect Concentration
NORMAN	Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances
P95-year	The 95th percentile concentration over the year
P99-year	The 99th percentile concentration over the year
PAF	Potentially Affected Fraction of species
PES	Probability of Effects on Species
PNEC	Predicted No Effect Concentration (protective benchmark used in EU-REACH)
PPDB	Pesticide Properties Database
REACH	EU regulation on Registration, Evaluation, Authorization and Restriction of Chemicals
SSD	Species Sensitivity Distribution
UFZ	Helmholtz-Zentrum für Umweltforschung
WFD	EU Water Framework Directive

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 2. Criteria for operationally characterizing ecotoxicity data as 'acute' (ECETOC 1993); longer exposure durations were classified as 'chronic'.

Species group	Acute test duration
Algae	12 h
Bacteria	12 h
Unicellular animals	12-24 h
Crustaceans	24-48 h
Fish	4-7 d
Mollusks, worms, etc	2-7 d

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 3. SSD-extrapolation scheme by factor From/To.

To / From	Order of extrapolation attempts to acute EC50[1]	Acute EC50 extrapolation factor[2]	Order of extrapolation attempts to chronic NOEC[1]	Chronic NOEC extrapolation factor[2]
Acute EC50	0	Multiply by 1	3	Multiply by 1/10
Acute NOEC	1	Multiply by 3	2	Multiply by 1/3
Chronic EC50	2	Multiply by 3	1	Multiply by 1/3
Chronic NOEC	3	Multiply by 10	0	Multiply by 1

[1] Numbers relate to quality scores in Table 4 [2] Numbers express inter-SSD extrapolations by parallel SSD shift.

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 4. Four-digit quality scoring of SSDs, based on underlying data types, quality and numbers.

Digit	Quality aspect	Modality	Meaning
1	SSD fullness	1	Data on full SSD available (Mu and Sigma)
1		2	Not sufficient data to calculate SSD slope
2	Biodiversity coverage	1	Number of taxa evaluated > 10
2		2	Number of taxa evaluated > 5
2		3	Number of taxa evaluated > 2
2		4	Number of taxa evaluated < 3
3	Data origin quality	1	Strictly measured
3		2	Extrapolated
3		3	Read-across
4	Extrapolation quality	1	Not extrapolated
4		2	Single-step extrapolation (Table 2)
4		3	Double-step extrapolation (Table 2)
4		4	Triple-step extrapolation (Table 2)
4		5	All available toxicity data extrapolation (Table 2)
4		6	Read-across

This article is protected by copyright. All rights reserved

Accepted Preprint

This article is protected by copyright. All rights reserved

Accepted Preprint

Table 5. Example of chemical ranking. An illustration of the top-15 ranked toxicants, their SSD-quality scores and their relative impact potential to species loss. Outcomes based on P95-msPAF-EC50, for Europe and for three example case study basins (relative ranks defined by Phorate for whole-Europe data defined as baseline ("1")). WFD- and NORMAN priority marks: compound listed as priority compounds for management attention following WFD or NORMAN methods. Note that the choice for P95 excludes peak exposures of e.g. pesticides (see SI-Section 6).

Substance	CAS	WFD priority	Norman priority	All 22 EU Basins	Danube basin	Rhine basin	Spanish basins	Count	Rank
Bisphenol-A	80-05-7	√	√	90316	85935	277239	110079	4	1
N-1,3-dimethylbutyl-N'-phenyl-p-phenylenediamine	793-24-8			2573	2451	20344	3528	4	2
Chlorpyrifos	2921-88-2	√		1629	2755		78285	3	3
Anthracene	120-12-7	√		1502	6675	3528	247	4	4
Octamethylcyclotetrasiloxane	556-67-2		√	1483	6325	1320	92	4	5
N-(4-aminophenyl)aniline	101-54-2			1381	268	15144	393	4	6
Cumene hydroperoxide	80-15-9			1123	4332	2242	634	4	7
Difenylamine	122-39-4		√	589	308	8667	1028	4	8
1-Dodecanol	112-53-8			48	558			2	9
Pyraclostrobin	175013-18-0			41	151	114		3	10
Cyhexatin	13121-70-5			25	8		2821	3	11
p-Phenylenediamine	106-50-3			19	17	97		3	12
Dimoxystrobin	149961-52-4			11	46	37		3	13
Terbufos	13071-79-9			6	75		96	3	14
Phorate	298-02-2			1			110	2	15
Count		3	3	15	14	10	11		
Sum relative score op 15 substances				100748	109904	328733	197314		

This article is protected by copyright. All rights reserved

Sum relative score remaining 1760 substances	208	571	0	0
Percent of sum relative score in top 15 substances	99.8%	99.5%	100.0%	100.0%

Accepted Preprint

This article is protected by copyright. All rights reserved

Accepted Preprint

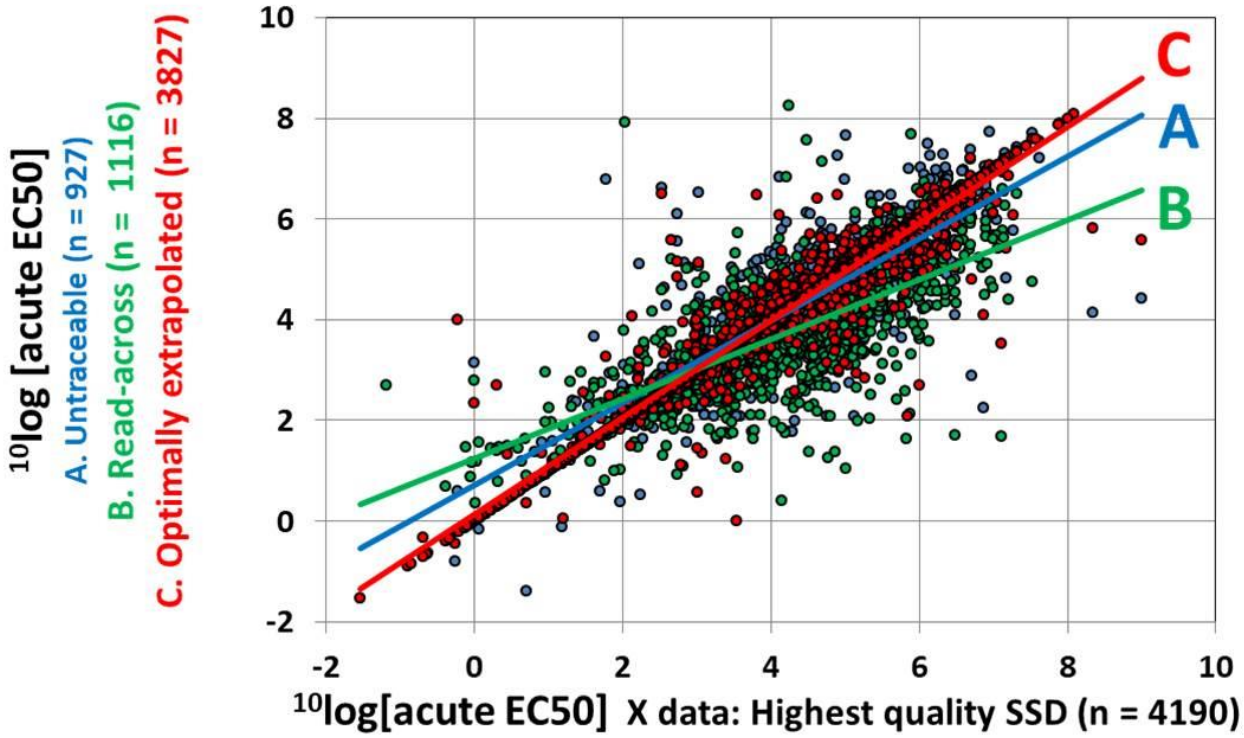

Figure 1

This article is protected by copyright. All rights reserved

Accepted Preprint

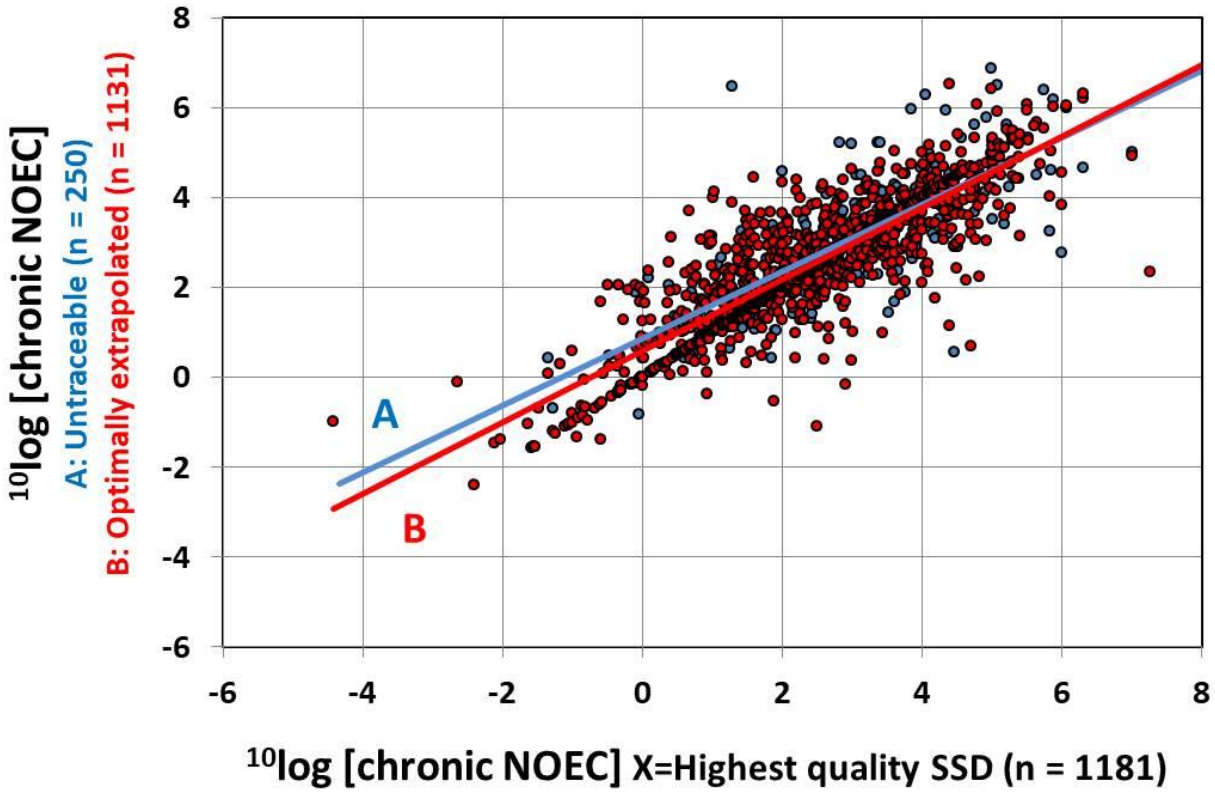

Figure 2

This article is protected by copyright. All rights reserved

Accepted Preprint

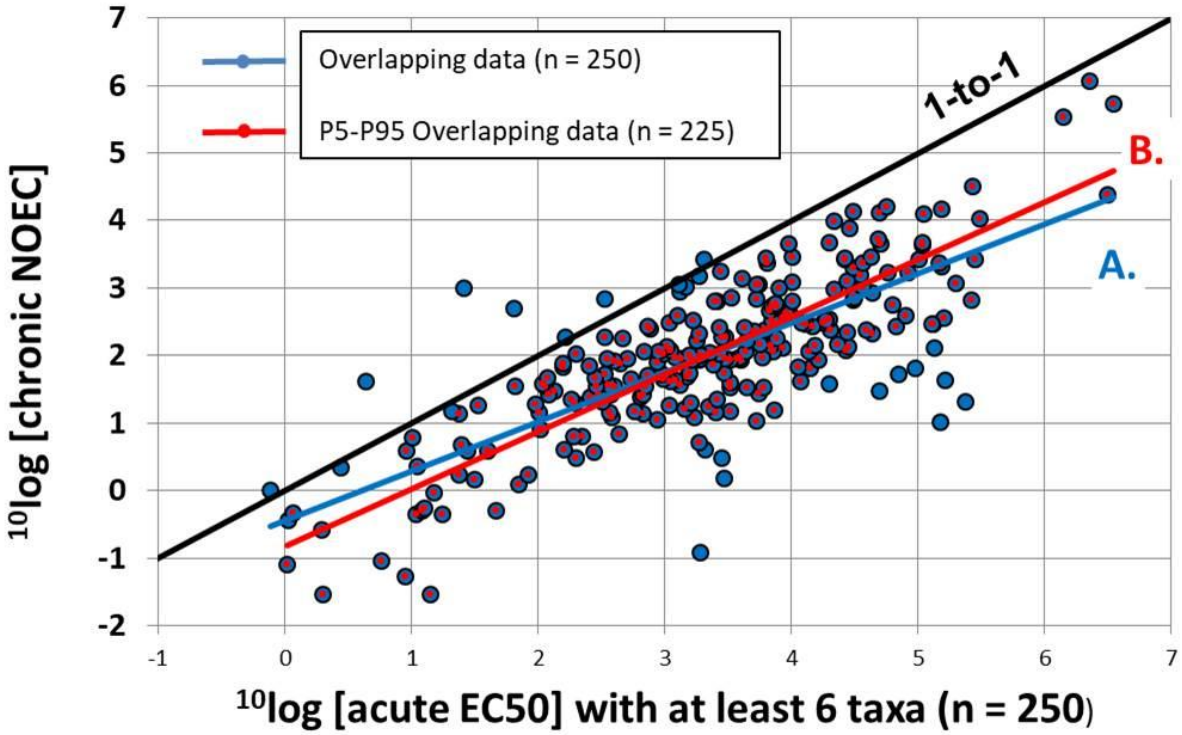

Figure 3

This article is protected by copyright. All rights reserved

Accepted Preprint

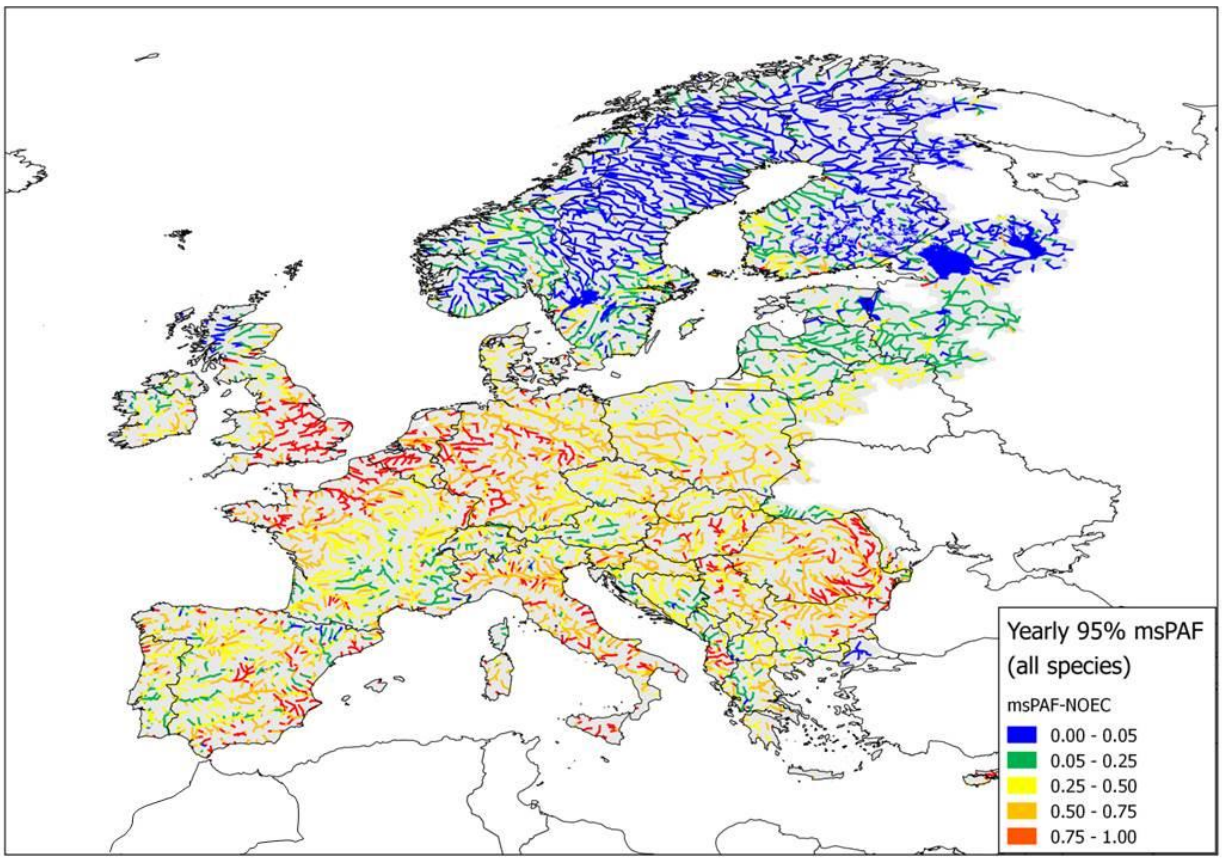

Figure 4

This article is protected by copyright. All rights reserved

Accepted Preprint

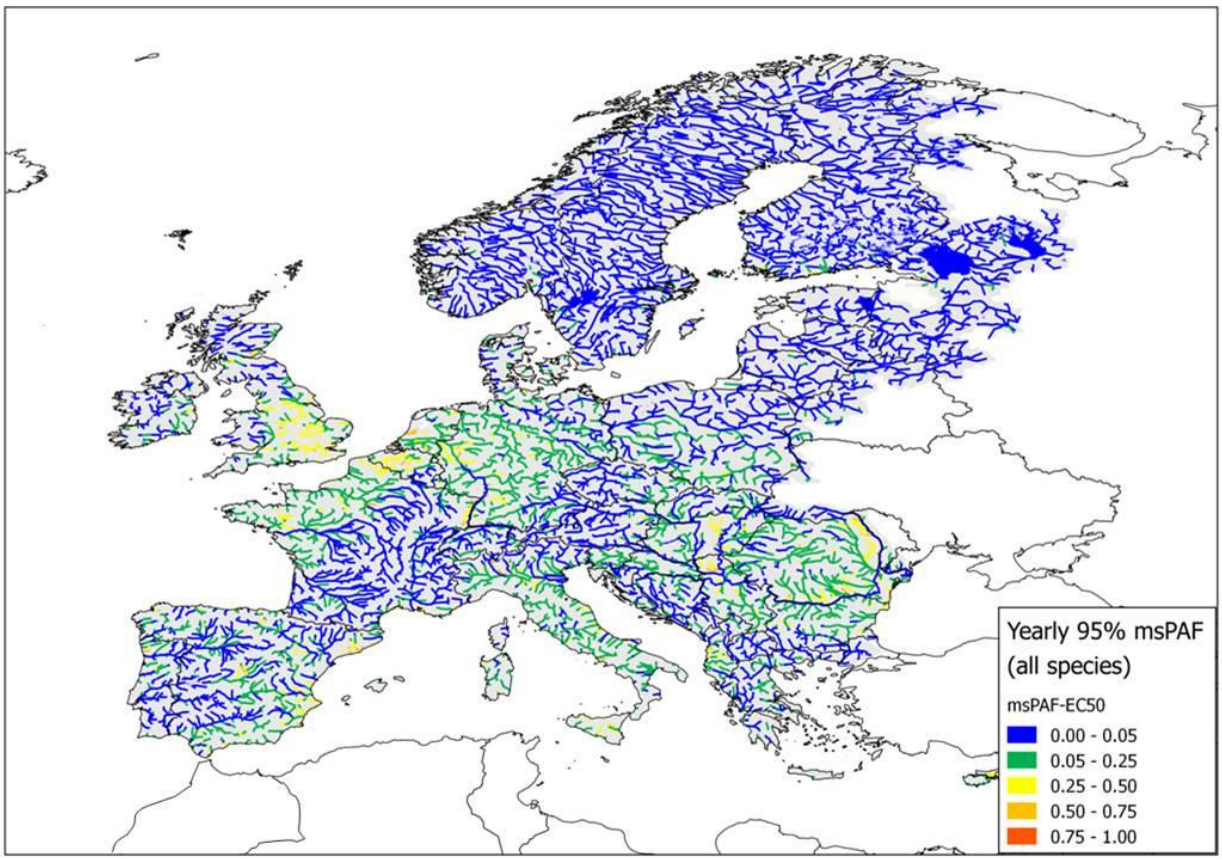

Figure 5

This article is protected by copyright. All rights reserved

Accepted Preprint

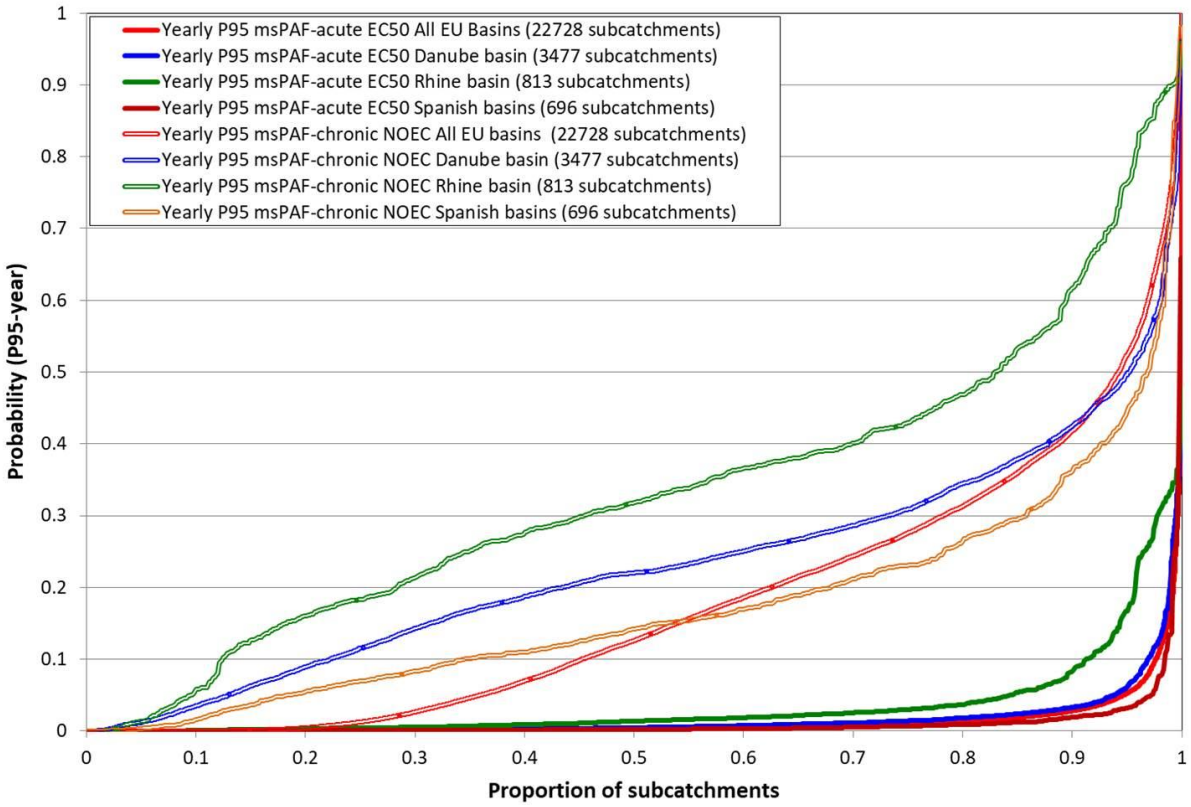

Figure 6

This article is protected by copyright. All rights reserved