COMPARISON OF MAF METHODS APPLIED TO ENVIRONMENTAL MONITORING DATA

29/11/2021

Karel Viaene

Charlotte Nys

Frederik Verdonck

ARCHE Consulting
Liefkensstraat 35D
B-9032 Wondelgem, Belgium

Contact
karel.viaene@arche-consulting.be
www.arche-consulting.be
Tel: + 32 9 216 70 34

PREFACE

This study was commissioned by the European Chemical Industry Council (CEFIC) and conducted by ARCHE Consulting.

On 14 October 2020 the EU Commission published its Chemical Strategy for Sustainability. This strategy calls to systematically integrate the issue of combined exposure (i.e. exposure to unintentional mixtures of different chemicals via our surroundings or the environment) into chemical risk assessments. This report focusses on reviewing of existing (monitoring) data and research with the aim of improving the scientific knowledge base of any future regulatory approach, with special focus on the methodology to determine a mixture assessment factor (MAF).

CEFIC has defined the scope and aims of the project, but ARCHE Consulting has performed all data collection, literature review and data analyses. Discussions were planned and held between the sponsor and the consultants during the course of the project. However, the content, analyses, discussion and conclusions presented in this report are the sole responsibilities of the authors. Any opinions and conclusions in this report are those of the authors and do not necessarily reflect those of CEFIC.

Project leaders were Karel Viaene and Frederik Verdonck at ARCHE Consulting and Steven Van de Broeck at CEFIC. For further information, please contact Steven Van de Broeck (sva@cefic.be) or Karel Viaene (karel.viaene@arche-consulting.be).

EXECUTIVE SUMMARY

A mixture assessment factor (MAF) to account for combination/mixture effects during the evaluation of risks for humans and the environment has been proposed by ECHA for integration in REACH. The current report does not discuss whether and how a MAF should be implemented, but focuses on, if a generic MAF would be implemented, how to best estimate such a generic MAF. MAF methodologies are developed based on a certain logic/reasoning and imply different assumptions. This impacts the outcome of MAF calculations and understanding these differences between MAF methodologies is key to a well-balanced MAF calculation. **Four different MAF calculation methods were evaluated** in this report: (1) equal toxic share of the most contributing substances (cfr. Van Broekhuizen et al., 2016); (2) the maximum cumulative ratio (MCR; Price & Han, 2011); (3) the size of the hazard index (HI); and (4) equal toxic share of all contributing substances (KEMI, 2021). The performance of these four MAF methods was evaluated with freshwater monitoring data. **Five environmental monitoring databases** were considered: one European-wide (Waterbase), three regional/national (Rhine, Adour-Garonne, Swedish Pesticides) and one local (Erft river) database.

The outcome of these MAF methods was **compared between databases and between methods**. Additionally, the **robustness** against the following criteria was evaluated: (1) impact of the choice of reference value (i.e. PNEC or HC5); (2) impact of the sample size and the presence of unknown and non-detected substances; (3) the consideration of mixtures with single substance risk and (4) the choice of protection level and level of conservatism.

Based on these criteria, the **maximum cumulative ratio was found to be most robust**: the choice of PNEC or HC5 had limited impact on the final MAF value, the MAF did not consistently increase for larger samples, calculated MAF values across databases were comparable (between 1 and 10) and differences between MAFs protective for 90% and 95% of the samples were limited (<1). The equal toxic share of the most contributing substance method was somewhat robust, but performed less well compared to the MCR for the sample size and level of protectiveness. The size of the hazard index was greatly affected by the choice of the reference value and sensitive to whether single substance risks were considered. The “equal toxic share of all contributing substances” method was also sensitive to the reference value used, the contribution of single substances and showed a linear increase of the MAF with increasing sample size.

Which method takes preference is partly a policy decision, but analyses like these help inform the decision-making. Methods differ mainly in how they distribute the burden of mixture toxicity: to all substances proportionally (e.g. maximum cumulative ratio) or disproportionally to the most contributing substances (e.g. equal toxic share of most contributing substances). Additionally, all these methods are very conservative, starting from worst-case assumptions (e.g. concentration addition for mixture toxicity) that can be further refined with higher-tier methods. This study confirms that differences between MAF methods can be considerable and future MAF discussions should take these differences into account.

ARCHE CONSULTING

TABLE OF CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

ARCHE conducted in 2021 a study on 'Characterising chemical co-exposures in EU to support a combined exposure assessment strategy'. The mixture pressure analysis showed that the vast majority of the observed mixture exposure are not at risk (81 – 94% of the observed monitored mixtures). For the remaining cases, it can be assessed which EU Regulation or Directive would be most effective to manage these mixtures of potential concern. The research done concludes that a potential mixture assessment factor (MAF) needs to be proportional to the magnitude of the mixture toxicity problem identified. The spatial use pattern analysis of observed mixtures in the surface waters indicated that the more downstream a river catchment, the more complex mixtures, the more likely higher number of chemicals (both in terms of number of chemicals as well as types of chemicals) will contribute to mixture toxicity and subsequently, a higher MAF, can be expected. However, since chemical oriented regulations are not always spatially explicit, a (high level) generic factor can be considered.

Various studies have suggested generic MAF factors. The most recent study is conducted on behalf of the Swedish Chemical Agency (KEMI) and suggests that a MAF should be in the order of magnitude between 10 and 50 (KEMI 2021). The current study provides an overview and comparison of the different possible methods to estimate MAF. In addition, the databases used in the two studies are tested to avoid potential discrepancies due to the databases used across both studies.

1.2. OBJECTIVES

Four objectives were defined for the current study.

1. Compare different MAF methodologies and understand how the choice of the methodology impacts the size of the MAF.
2. Assess how the database affects the size of the MAF i.e. how different are MAF values between different databases and what are the implications.
3. Assess the impact of the used reference values to identify key mixtures contributors.
4. Assess the impact of the number of chemicals on the MAF size.

2. METHODOLOGY

2.1. MAF METHODOLOGY

Four different methods to derive a potential Mixture Assessment Factor (MAF) based on risk calculations are discussed in the present study. All methods depart from the Concentration Addition-based Hazard Index/Hazard Quotient approach, but different outputs of this approach are used to derive the MAF in the various methodologies. Central to the four methods is the mixture risk estimation that is expressed as the **Hazard Index** (Eq. 1):

$$HI = \sum HQ = \sum \frac{c_i}{RV_i} \qquad \text{(Eq. 1)}$$

In Equation 1, HI is the Hazard Index (in the CEFIC-MIAT terminology) which represents the cumulative risk ratio (or risk quotient used as terminology by other authors). The Hazard Index is the sum of the Hazard Quotient of all measured substances in the mixture sample. The Hazard Quotient of a substance *i* is defined by the ratio between the measured concentration of substance *i* (c_i) in the mixture sample and the Reference Value of substance *i* (RV_i). The Reference Value can be any environmental threshold concentration, but for the purpose of this study, the chronic 5% hazardous concentration (HC5) has been selected as default for mixture pressure calculations (see Section 2.2). However, also an analysis of the influence of using the Predicted No-Effect Concentration (PNEC) vs. the chronic HC5 on the MAF derivation methodologies has been performed (see Section 2.3).

2.1.1. EQUAL TOXIC SHARE OF MOST CONTRIBUTING SUBSTANCES

In a first method, it is assumed that the carrying capacity in the environment is to be equally split across the most contributing substances. **The size of the MAF can be directly linked to the number of substances contributing to the mixture cumulative risk**. The larger the number of contributing substances, the larger the MAF factor.

Assuming a mixture of n substances and a worst case of co-existence at equal toxic shares of 1/n of their individual PNEC, would lead to the derivation of a preliminary MAF of n. A similar rationale is presented in Van Broekhuizen et al. (2016). However, in the present study, a reasonable worst-case *n* is derived for each database as the number of substances that contribute to at least 90% of the mixture pressure (defined as the Hazard Index) for approximately 90% of the mixture samples.

2.1.2. MAXIMUM CUMULATIVE RATIO

The second method relies on the Maximum Cumulative Ratio (MCR) to derive a MAF. The MCR has been defined in the CEFIC-MIAT methodology as **the ratio of the Hazard Index to the Hazard Quotient (HQ) of the compound with the highest Hazard Quotient** (Price and Han 2011; Price et al. 2012a; Eq. 2).

$$MCR = \frac{HI}{Max\ HQ} \qquad \text{(Eq. 2)}$$

The smaller the MCR, the smaller the number of compounds that are significantly contributing to the overall mixture risk. Using the MCR, the "nature of concern" of a mixture and related risk management actions can be decided on. This way, the MCR could equally be considered as a surrogate for the MAF. The MCR does not assume equal toxic share of most contributing substances in a mixture (as in previous section) but it does proportionally distribute MAF burden across all substances.

2.1.3. SIZE OF HAZARD INDEX

The third method relates the MAF to the Hazard Index (HI; Eq. 1), which represents the cumulative (potential) risk of a mixture. **The size of the HI can be considered as the factor required to turn a potential concern conclusions (HI >1) into no risk conclusion (HI <1).**

For the derivation of the MAF based on the Hazard Index, the following mixture groups are defined based on the HI and MCR (based on Price et al. 2012):

- Group I are the combined exposures that are a potential concern because one or more individual chemicals are a concern (i.e. at least one of the HQ_i>1).
- Group II are the combined exposures where there is a low concern for both individual chemicals and for their combined effects: (i.e. HI<1).
- Group III are the combined exposures where there is a low concern for individual chemicals but there is a potential concern for the combined effects (HQ_i<1, HI>1).

A high-end percentile has been considered as reasonable worst-case for derivation of a MAF. The percentile were calculated for all Groups, for Group III only and for Group III + Group I managed (in which HQ > 1 for individual substances were capped at 1).

2.1.4. EQUAL TOXIC SHARE OF ALL CONTRIBUTING SUBSTANCES

This algorithm to calculate the MAF is described in detail in the report of KEMI (2021), and is **based on the concentration addition-based Hazard Index/Hazard Quotient approach**. It should be noted that the

terminology used in KEMI (2021) to describe toxic pressure is slightly different compared to those used in the previous sections, but these are equivalent. The risk quotient (RQ) used in KEMI (2021) is equivalent to the Hazard Quotient (HQ), while the sum of risk quotients (RQ_{sum}) is equivalent to the HI in Section 2.3.

In short, the algorithm contains three major phases. In **Phase I** concentrations are taken at face value and the RQ_{sum} is calculated. If the RQ_{sum} <1 for a mixture sample, no MAF is needed and the algorithm stops. If RQ_{sum}>1, phase II of the algorithm starts.

In **Phase II** single-substance exceedance of the reference value are considered by setting the RQ of substances for which the measured concentration exceeds their reference value to 1 (referred to as the managed RQ: RQ_{man}=1). Using these RQ_{man}, the $RQSum_{Man}$ is calculated for the entire sample. Although by capping the risk quotient of each individual mixture component at a value of 1.0 the resulting mixture conforms with existing regulations on the registration and authorization of individual chemicals, $RQSum_{Man}$>1 and therefore potential risks due to mixture exposure are still predicted and the sample should be considered in Phase III.

Phase III consists of the MAF calculation in an iterative manner. In a first step, a MAF of *n* (with *n* the number of mixture constituents) is applied. In practice, only those RQ_{MAF} values that exceed 1/MAF are affected by the MAF, and are set to the RQ_{MAF} value of the most contributing substance. Other RQ are kept at the original value from Phase II. The $RQSum_{MAF}$ is calculated. This lowers the $RQSum_{MAF}$ almost always to a value <1, leaving room for a less stringent MAF. After this first step, the MAF is iteratively lowered until $RQSum_{MAF}$ equals exactly 1.

The application of the MAF algorithm on an example mixture is illustrated in Table 1 of the report of KEMI (2021). Because the RQ_{MAF} of substance for which the RQ_{MAF} exceeds 1/MAF is set to the RQ_{MAF} value of the most contributing substance, the reduction factor for each of the substances in the mixture is different. While small contributors are not affected (i.e. if RQ< 1/MAF) in the algorithm, the most contributing substance is most affected with a reduction factor equal to the MAF. The substance, that has a RQ between the RQ of the highest contributing substance and 1/MAF, has a reduction factor that equals the original RQ of the substance/RQ_{MAF} of the most contributing substance.

2.2. CALCULATION OF 5% HAZARDOUS CONCENTRATIONS

The basis of calculations of mixture pressure in the current study is the 5% hazard concentration (HC5), because these values are more robust compared to PNEC values where assessment/safety factors are included. HC_5 values were calculated based on log chronic NOEC/EC_{10} and the slope of the species sensitivity distribution (SSD) collated by Posthuma et al. (2019). The dataset of Posthuma et al. (2019) consists of parameters (median and standard deviation of log-transformed toxicity data) of log-normal SSDs for 12 386 chemicals. Separate parameters are published for both the acute and chronic SSD. Toxicity data for the derivation of the SSDs was collated from different sources, such as the US EPA's ECOTOX database, REACH data. For substances for which insufficient toxicity data could be extracted, Posthuma et al. (2019) used read-across approaches to estimate the remaining SSD-parameters. SSDs parameters were derived when toxicity data for at least 3 species was retrieved. Remaining data gaps between acute and chronic SSDs were tackled with extrapolation factors (Posthuma et al. 2019). The necessary data were extracted from the Posthuma et al. (2019) database based on CAS-numbers. If an exact match was not found, the search was repeated based on the substance name. For a few substances, no match based on CAS-number of substance name could be found, for these substances an environmental threshold level was derived from the WFD-directive (i.e. annual average-environmental quality standard (AA-EQS), EC 2013/39/EU), from the PNEC reported by Gustavsson et al. (2017) or Markert et al. (2020) or PNECs, RACs and EQS values for newer pesticides, retrieved from the ETOX database of the German Environment Agency (http://webetox.uba.de/webETOX/index.do).

The calculated chronic HC5 represent different data-qualities. Table 1 shows the different quality categories for the chronic SSDs selected from Posthuma et al. (2019). For the Rhine dataset (see section 3.4), 67% of the chronic SSDs were not extrapolated (i.e. they were derived based on chronic toxicity data). 29% of the chronic SSDs were extrapolated based on acute EC_{50} SSD and 4% based on the acute NOEC SSD. For the Adour-Garonne dataset (see section 3.3), 51% of the chronic SSDs were not extrapolated (i.e. they were derived based on chronic toxicity data). 29% of the chronic SSDs were extrapolated based on acute EC_{50} SSD and 7% based on the acute NOEC SSD. For 10% of the substances, the chronic SSD was extrapolated from a poorly presented Acute SSD, implying that the acute SSD was derived from read across or that the acute SSD contained only 1 toxicity data point. For the remaining substances, there was no match found within the Posthuma database, for these data other sources were used to obtain an effect threshold.

Table 1 Data quality of SSD parameters extracted from the database of Posthuma et al. (2019) for the Adour Garonne and Rhine dataset. Numbers indicate number of substances belonging to each category

SSD extrapolation category	Number of species in SSD	Rhine	Adour-Garonne	Swedish pesticides	Erft
Chronic NOEC not extrapolated	Officially enough species (>10) for ERA with SSDs[a]	34	31	38	27
	Enough species (6-10) for ERA with SSDs	-	1	6	3
	Marginally enough species (3-5) for ERA with SSDs	-	2	1	-
Chronic NOEC extrapolated from Acute EC50[b]	Officially enough species (>10) for ERA with SSDs	12	11	16	17
	Enough species (6-10) for ERA with SSDs	3	4	11	6
	Marginally enough species (3-5) for ERA with SSDs	-	7	5	7
Chronic NOEC extrapolated from Acute NOEC[c]	Officially enough species (>10) for ERA with SSDs	2	-	17	8
	Enough species (6-10) for ERA with SSDs	-	-	5	3
	Marginally enough species (3-5) for ERA with SSDs	-	3	2	3
Chronic NOEC extrapolated from poorly presented Acute SSD[d]	Acute EC50 for 1 species	-	1	-	
	Read across	-	7	5	15
Other sources: WFD-EQS, Restriction dossier, member state EQS		-	3	5	4
Other: Target Lipid Model PNEC (PAH only)[e]			7		
Total number of substances considered in cumulative risk estimation		51	76	107	93
Number of substances for which no threshold value could be derived		0	0	3	5

[a] 'Officially enough species' refers to the minimal sample size put forward for the use of statistical derivation methods as described in the Technical Guidance Document on Risk Assessment (EC 2003).

[b] An extrapolation factor of 10 was applied by Posthuma et al. (2019) to extrapolate from acute EC50 to chronic NOEC, i.e. the acute EC50s were divided by 10.

[C] An extrapolation factor of 9 was applied by Posthuma et al. (2019) to extrapolate from acute NOEC to chronic NOEC, i.e. the acute NOECs were divided by 9.

[d] An extrapolation factor of 10 was applied by Posthuma et al. (2017) on the median acute EC50 calculated from read-across or the poorly represented acute SSD, and a slope of 0.7 was assumed (Posthuma et al. 2019).

[e]For PAH, chronic PNECs were extracted from the PETROTOX-tool, which calculates calculated PNECs based on the Target Lipid Model (Redman et al. 2017).

2.3. COMPARISON OF THE USE OF HC5 VS. PNEC AS REFERENCE VALUE

For two databases (i.e. the Swedish Pesticides and the Erft river, see section 3), the influence of the selected reference value, HC5 vs. PNEC values on the calculation of the MAF techniques is evaluated. For this evaluation, PNECs were taken directly from the respective studies. For the Swedish Pesticides, environmental threshold values were taken from the study by Gustavsson et al. (2017). The PNECs published by Gustavsson et al (2017) represent Swedish Water Quality Objectives (WQO's). WQO's reported by Gustavsson et al. (2017) were derived using a method that closely follows the REACH approach for deriving Predicted No Effect Concentrations (PNEC) values, based on single species data and assessment factors between 10 and 1000, depending on the underlying ecotoxicological endpoints. For the Erft river, 'assessment values' were taken from the study by Markert et al. (2020). The 'assessment values' published by Markert et al. (2017) represent regulatory established environmental quality standards (EQS), such as the those from the WFD or related national legislation on water quality. In the absence of EQS or if new ecotoxicological data have recently been published, Markert et al. (2020) derived assessment values from validated ecotoxicological data, i.e. PNEC, EQS proposals or threshold values for national monitoring programs. The PNEC values from the studies of Gustavsson et al (2017) and Markert et al. (2020) were amended by PNECs, RACs and EQS values for newer pesticides, retrieved from the ETOX database of the German Environment Agency (http://webetox.uba.de/webETOX/index.do).

3. DESCRIPTON OF DATABASES AND DATA PROCESSING

3.1. OVERVIEW OF DATABASES AND THEIR GENERAL TOXIC PRESSURE

Five monitoring databases were covered in the current study: the European-wide Waterbase, the ICPR Rhine database, the EauFrance database for Adour-Garonne, the Swedish pesticides database and the monitoring campaign in the river Erft (Table 2). All databases are available either online or on request at the data-owners. All databases are discussed below and in ARCHE and VITO (2021) for further details.

Table 2: Overview databases covered in this study.

Database	Number of samples	Number of analytes
Waterbase	12356	334
Rhine	394	57
Adour-Garonne	3557	77
Swedish pesticides (Kemi 2021 study)	1513	109
Swedish pesticides (current study)	1891	109
River Erft	503	153

Figure 1 shows the general toxic pressure in the different datasets. The Adour-Garonne dataset represents generally a low toxic pressure, with 94% of the mixture samples predicted to show no (potential) risk (i.e. HI<1,). The Rhine and the Swedish dataset represent moderate toxic pressure with more than 10% of the samples predicted to be at risk mainly due to single substances (at least one substance HQ>1; Sweden) or mixture exposure (HI>1; Rhine). The Erft river shows a high toxic pressure, but mainly due to single substances, as in almost 50% of the samples at least one substance HQ>1.

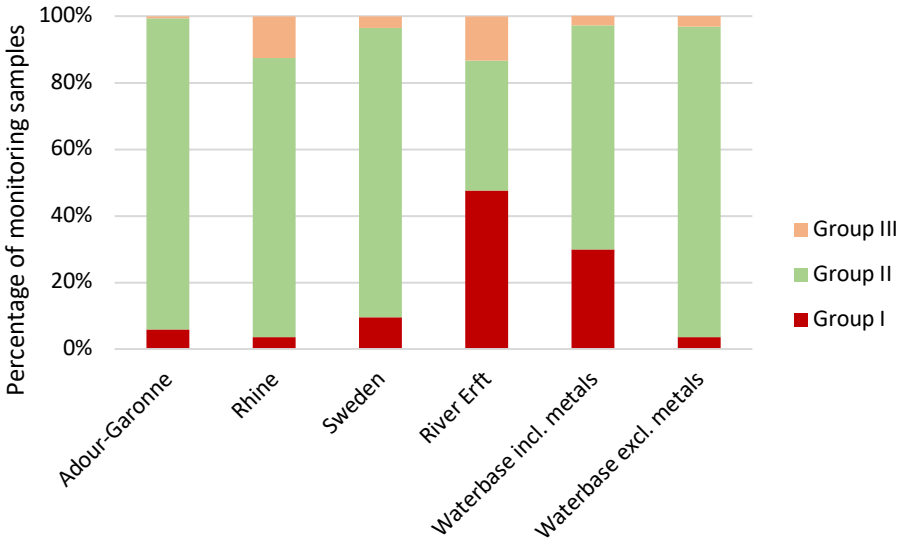

Figure 1: Toxic pressure of mixture samples in the different databases represented by the percentage of monitoring samples belonging to three different mixture groups. Group I are the combined exposures that are a potential concern because one or more individual chemicals are a concern (i.e. at least one of the HQ_i>1). Group II are the combined exposures where there is a low concern for both individual chemicals and for their combined effects: (i.e. HI<1). Group III are the combined exposures where there is a low concern for individual chemicals but there is a potential concern for the combined effects (HQ_i<1, HI>1). Toxic pressure was calculated using the HC5 as reference value. Calculations for the Waterbase dataset were obtained from Rodea- Palmorales et al. (submitted).

3.2. WATERBASE

Waterbase is the generic name given to the EEA's databases (European Environment Agency) on the status and quality of Europe's rivers, lakes, groundwater bodies and transitional, coastal and marine waters, on the quantity of Europe's water resources, and on the emissions to surface waters from point and diffuse sources of pollution. The dataset contains time series of nutrients, organic matter, hazardous substances and other chemicals in rivers, lakes, groundwater, transitional, coastal and marine waters. The data has been compiled and processed by EEA.

For this project, the database on chemical status and quality was used (Waterbase quality ICM). Specifically, we used the dataset T_WISE6_DisaggregatedData containing the disaggregated water quality data on the observed values (e.g. concentrations) of determinants in water, sediment and biota samples in inland, coastal and marine waters as reported by EEA Member Countries on an annual basis.

This study used the Waterbase dataset as processed by Rodea-Palomares et al. submitted. Yearly mean and maximum values were calculated per site. The selected reference values for mixture pressure assessment by Rodea-Palmorales et al. (submitted) were the chronic HC5 derived based on the SSD-parameters published by Posthuma et al. (2019). The only exception were polyaromatic hydrocarbons for which the PNEC reported in the PETROTOX tool were used (Redman et al. 2017). Metals were not included in the analysis because these are naturally occurring substances where background concentration and bioavailability are critical to derive an accurate estimation of the risks. However, it was not possible to perform for these corrections due to a lack of data and resources, hence metals were excluded from the analysis.

A database with HQ for each substance and per site was taken from the supplementary information from Rodea-Palmorales et al., submitted.

3.3. ADOUR-GARONNE (EAUFRANCE)

In France, there are 6 water agencies established since 1964 consisting of 6 big water basins Adour-Garonne, Loire-Bretagne, Seine-Normandie, Artois-Picardie, Rhin-Meuse and Rhône-Méditerranée. These water agencies carry out a mission for general interest, including managing, preserving water sources and aquatic environment. The analysis on water quality has been started by these agencies ever since. The data is annually updated in the database of eauFrance and can be downloaded at http://www.naiades.eaufrance.fr/.

In scope of this study, a local dataset of Adour-Garonne of period 2015-2019 was used. This dataset can be downloaded separately at this link. The measurement was started in 1971 and has been regularly updated. For the current study, period 2015-2019 was selected. The data package includes physico-chemistry, phytosanitary and hydrobiology data.

The physico-chemistry data includes more than 2000 sampling sites covering 7 sub basins in the region, with more than 409 variables. The focus is set on chemical substances which are consisting of 47% industrial substances, 12% agrochemical substances, 21% pharmaceutical substances, 5% PAHs and 15% of mixed-use substances.

3.4. RHINE (ICPR)

The state of the Rhine is being monitored from Switzerland until the Netherlands by ICPR – International Commission for the Protection of the Rhine. The first monitoring was carried out as early as the 1950ies, so that, for certain substances in the Rhine, corresponding time series are available. Over time, more and more

substances were included in monitoring activities, the measurement of suspended matter was expanded and the sampling frequency was increased. Monitoring results are supplemented every year, thus continuing time series. The data can be downloaded at https://www.iksr.org/en/topics/water-quality/water-quality-data.

The monitored parameters include both organic micropollutants, inorganic compounds and general water quality parameters. The monitored chemicals are mainly priority substances under the EU Water Framework Directive. The sampling intensity has increased over time. The monitoring locations are mainly located at the river Rhine and some major tributaries. The number of monitoring locations is rather small. Only about 6 monitoring locations have been intensively monitored.

The version of the database used in the ARCHE and VITO 2021 study was used here (ARCHE and VITO 2021). The data was downloaded on February 11, 2021. The data have been processed to maximize both the number of chemicals per sample and the total number of samples. The final dataset contains 394 samples where all 57 analytes have been measured but not necessarily detected. Reference values for calculation of toxic pressure are the chronic HC5 as discussed Section 2.5, except for Polyaromatic Hydrocarbons (PAH) for which the chronic PNEC calculated from the Target Lipid Model was used (Redman et al. 2017, see ARCHE & VITO 2021). It has been argued that some of the risk evaluation methods, such as the HI/HQ approach, are not suitable for naturally occurring substances, because they might predict risks below natural background concentrations (Van Regenmortel et al. 2017, Nys et al. 2018). Therefore, to account for background concentrations, the added concentration approach has been used for naturally occurring substances. The added (risk) approach assumes that species are fully adapted to the natural background concentration and therefore contributes to the toxicity. This approach accounts for background concentrations of naturally occurring substances by subtracting the natural background from the measured environmental concentration. Natural background concentrations of metals in European waters were considered to be the lower 10th percentiles of metal concentration in the geochemical baseline FOREGS database (Table 3).

Table 3 Overview of background concentrations[a] of metals in European surface waters.

Metal	Background concentration (µg/L)
Cd	0.01
Cr	0.38
Pb	0.09
Cu	0.88
Zn	2.68
Ni	1.91

[a] Background concentrations are represented here as the lower 10th percentile of metal concentrations in the FOREGS database

3.6. SWEDISH PESTICIDES

Sweden is regularly monitoring pesticide concentrations in four agriculturally dominated catchments and two rivers since 2002. These data are available for download from http://jordbruksvatten.slu.se/. An extended analysis of het mixture risk of these pesticides has been performed by Kemi in 2021 (KEMI 2021). A similar approach to collect and process the data was followed in the current study. The data was downloaded on October 12, 2021. All datapoints prior to (but excluding) 2019 were selected to attain a comparable dataset to the 2021 Kemi study. The dataset contained 1891 samples for 109 pesticides. All values below the analytical level of quantification were set to zero prior to the analysis.

3.7. ERFT RIVER

The Erft river is located in the South-west of Nordrhein-Westfalen, Germany (Markert et al. 2020). The river is 104 km long and has a catchment of 1918 km^2. The middle and lower reaches of its course are strongly influenced by intense agriculture, urban and industrial areas. The Erft catchment is marked by a high population density of 665 inhabitants/km^2. Additionally, active and former lignite mines are present in the catchment area. This translates to exposure to a wide variety of chemicals. Samples were taken at 39 sampling sites along the Erft river and seven tributaries over 13 sampling campaigns (March 2016 - March 2017).

The data for the Erft River was discussed in the Kemi 2021 report and has been previously used in the study by Markert et al. (2020). Over 150 chemicals were monitored in the German river Erft between 2016 and 2017, including pesticides, industrial chemicals, personal care products, pharmaceuticals and their degradation products (Markert et al. 2020). The final dataset consists of 503 samples and 98 detected chemicals. All monitoring data and environmental threshold values were kindly provided by Dr. Markert (German North Rhine Westphalian State Agency for Nature, Environment and Consumer Protection; LANUV) and the Erftverband. Concentrations detected at levels below the analytical level of quantification were set to zero. Otherwise, the reported data have been taken at face value, without further assessment or modification.

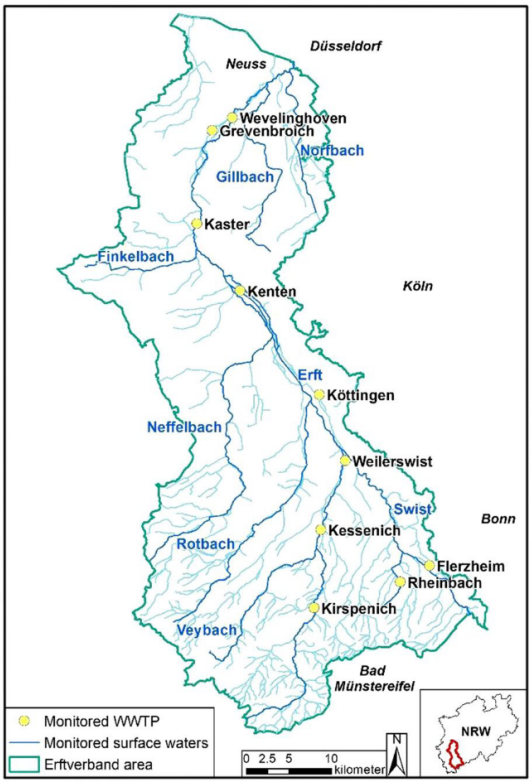

Figure 2: The Erft river catchment area and monitoring stations. (Source: Markert et al. 2020)

4. COMPARISON OF MAF SIZE FOR EACH MAF METHOD

4.1. EQUAL TOXIC SHARE OF MOST CONTRIBUTING SUBSTANCES

The different analysed data sets and the scientific literature provide an indication on the number of substances contributing to mixture toxicity. The contribution of the risk drivers in the Adour-Garonne, Rhine, Swedish Pesticides and Erft river database was calculated and is visualized in Figure 3 and Figure 4. For the Adour-Garonne dataset, one substance explains 95% of the hazard index for 50% of the mixture samples and the first three risk drivers explain 90% of the hazard index for 94% of the mixture samples. For the Rhine dataset (being a data set of hot spots), two substances explain 92% of the hazard index for half of the mixture samples and the first six risk drivers explain 90% of the hazard index for ca. 92% of the mixtures. For the Swedish Pesticides dataset, three substances explain 95% of the hazard index for half of the mixture samples and the first five risk drivers explain 90% of the hazard index for ca. 92% of the mixtures. For the Erft dataset, three substances explain 95% of the hazard index for half of the mixture samples and the first four risk drivers explain 90% of the hazard index for ca. 94% of the mixtures.

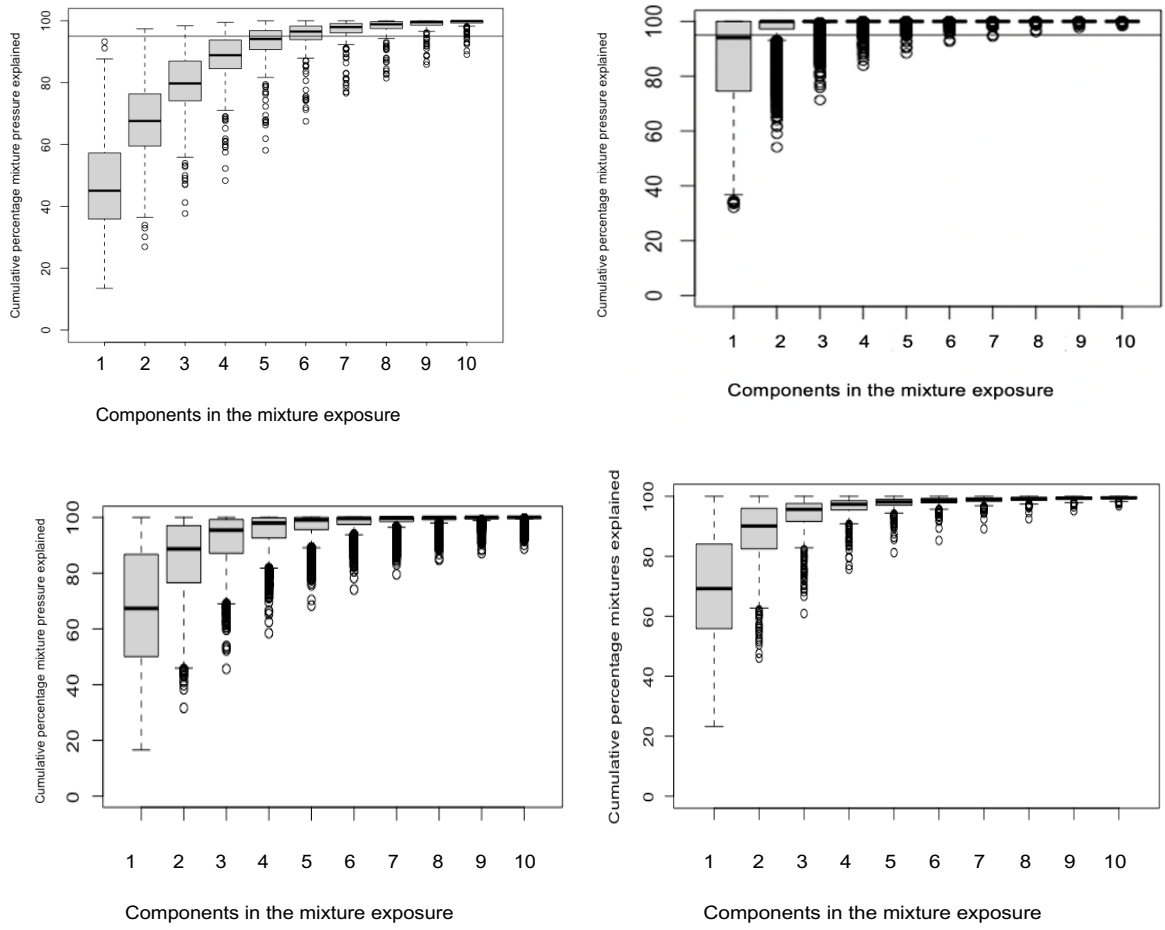

Figure 3: Cumulative percentage of mixture pressure explained by the different mixture components for all exposures for the Adour-Garonne dataset (upper left panel), Rhine dataset (upper right panel), the Swedish Pesticides dataset (lower left panel) and the Erft dataset (lower right panel). Mixture pressure is expressed as Hazard Index (based on HC5).

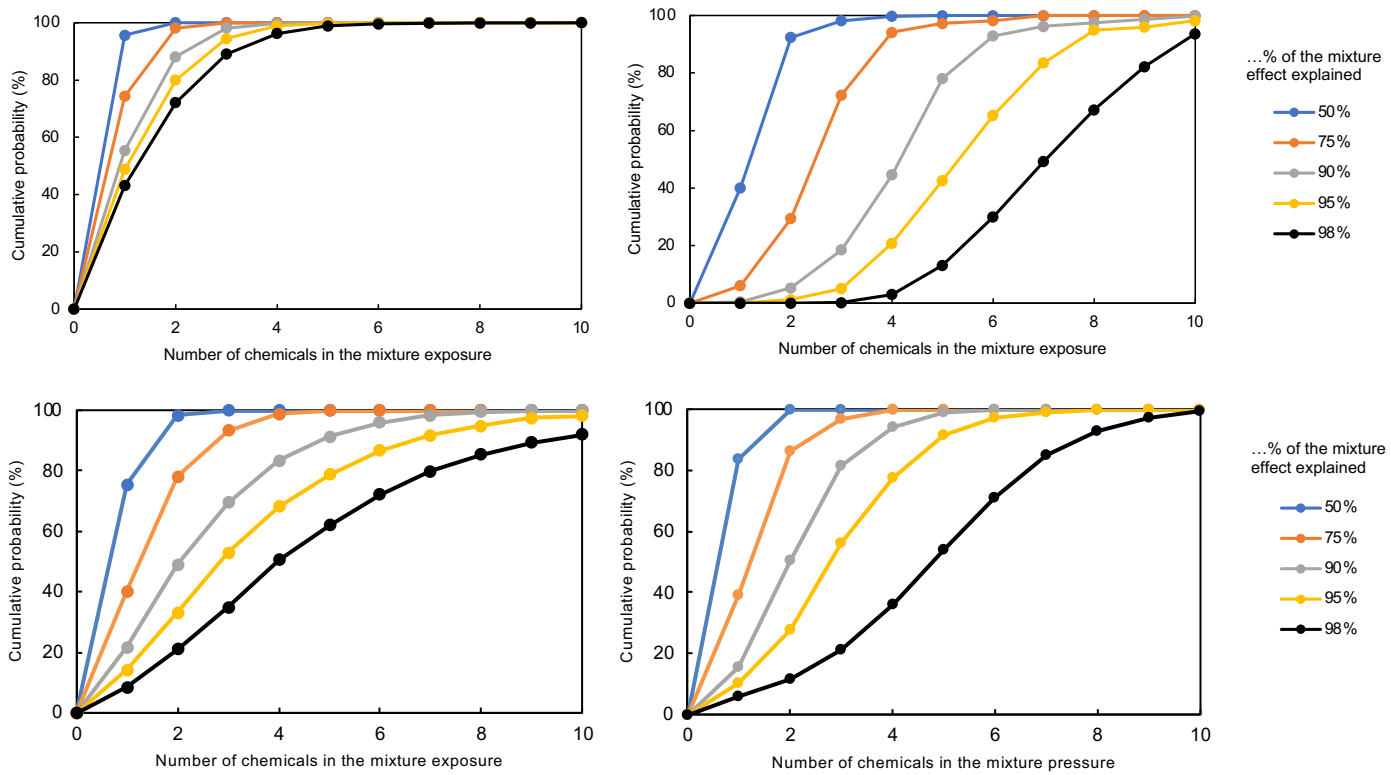

Figure 4: Cumulative probability in function of chemicals contributing for different percentages to the total mixture pressure for the Adour-Garonne dataset (upper left panel), Rhine dataset (upper right panel), the Swedish Pesticides dataset (lower left panel) and the Erft dataset (lower right panel). Mixture pressure is expressed as Hazard Index (based on HC5).

Rodea-Palomares et al. (submitted) calculated the contribution of first three risk drivers to overall Hazard Index (HI, i.e. cumulative risk) for the EU Waterbase (see Figure 5). Overall, one substance explains 75% of the hazard index for 50% of the mixture samples and the first three risk drivers explain 90% of the hazard index for 65% of the mixture samples. Nine substances are needed to explain 90% of the hazard index in 90% of the mixtures . This number is higher compared to other data sets. Note that Waterbase also contains a larger number of substances.

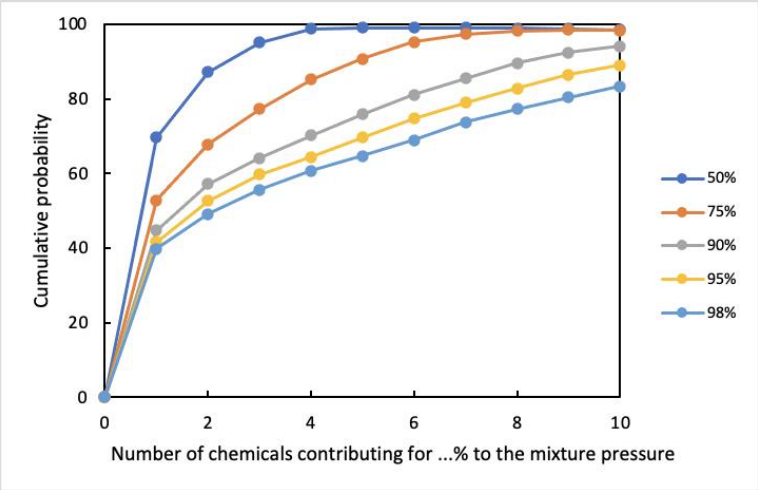

Figure 5: Cumulative probability in function of chemicals contributing for different percentages to the total mixture pressure (based on data from Rodea-Palomares et al., 2021).

Several other studies in the scientific literature show that usually risks from chemical mixtures are driven by a limited number of mixture components:

- Verro et al. (2009) found that in the river Meolo in Italy one or a few compounds were usually responsible for more than 80% of the CA estimated mixture toxicity.
- Backhaus and Karlsson (2014) investigated environmental risks from pharmaceuticals detected in waste water treatment plant effluents across Europe and identified that at maximum 10 compounds usually explained more than 95% of the overall risk. Note that two compounds failed single substance risk assessment in the seven samples assessed.
- Posthuma et al. (2016) found for Dutch rivers that 5–10 compounds are usually responsible for nearly the whole toxic pressure. The compounds driving the risk from the chemical mixtures might differ depending on the specific site/mixture composition, but the number of compounds responsible for the biggest fraction of the mixture effect seems to be always in the same range. Based on this, Van Broekhuizen et al. (2016) suggested about 5-10 chemicals dominate in cumulative risk to account for >95% of the combined effect, i.e. a coverage of >95% while recognizing the need for a broader EU study.
- Vallotton and her co-workers analyzed the joint toxicity of pesticide mixtures detected in the water quality monitoring of the National Water-Quality Assessment (NAWQA) program of the U.S. Geological Survey (Vallotton et al., 2016). They concluded that the environmental risk for more than 90% of the samples analyzed is driven by just one compound.
- Gustavsson et al. (2017) reported that pesticide mixture risks are often driven by only 1-3 compounds. However, the risk-drivers (i.e., individual pesticides explaining the largest share of potential effects) differ substantially between sites and samples. Gustavsson et al. (2017) also indicated that the risk drivers are quite specific for each sample and site: site E21, for example, has a median MCR for the

RQWQO of 2.9, indicating that only between 2 and 3 compounds dominate the estimated mixture risk. But in order to capture at least 95% of the overall risk of all 248 samples taken over the years, there are still 44 compounds that need to be monitored and the situation at the other monitored sites is very similar. This is due to the fact that the dominating compounds constantly fluctuate between samples.

The size of the MAF can be directly linked to the number of substances contributing to the mixture cumulative risk. The larger the number of contributing substances, the larger the MAF factor. Assuming a mixture of n substances and a worst case of co- existence at equal toxic shares of 1/n of their individual PNEC, would lead to the derivation of a preliminary MAF of n. A similar rationale is presented in Van Broekhuizen et al. (2016).

Based on the findings in literature and our own dataset analysis, one can conclude that, assuming that the number of equal toxic share of most contributing substances in a mixture is a proxy for a MAF, for broad coverage databases and the majority (ca. 90%) of mixtures, 3-8 substances are contributing to ca. 90% of the hazard index or cumulative risk, suggesting an MAF between 3 and 8.

4.2. MAXIMUM CUMULATIVE RATIO

Table 4 provides the 90th and 95th percentiles and the maximum MCR values for selected representative data sets. The 90th percentile is between 1.7 and 4.2 whereas the maximum varies between 3 and 11.

Table 4: MCR or Maximum Cumulative Ratio[a] for representative data sets based on lower tier assessment (based on more realistic scenario in which non-detects are excluded).

		90th percentile MCR	95th percentile MCR	Maximum MCR
Current study	Adour-Garonne	1.7	2.0	3.1
	Rhine	3.3	3.7	4.4
	Swedish Pesticides	2.5	2.9	5.2
	Erft	2.3	2.7	4.3
Rodea-Palomares et al. (in preparation)	EU Waterbase	4.2	5.1	11

[a] MCR represents the Hazard Index divided by the highest Hazard Quotient

Gustavsson et al. (2014) reported median MCR of 2.65 over alle sites using the PNEC as reference value (median MCR ranging between 2.0 and 3.2 over the different sites). A review of case studies (Bopp et al. 2016) showed that the MCR was usually ranging between 1 and 7, depending on the number of analysed compounds (Price and Han 2011; Backhaus and Karlsson 2014; De Brouwere et al. 2014; Price et al. 2012, 2014, Valloton & Price 2016, Gustavsson et al. 2017), that is, the number of chemicals that drive the mixture risk is usually low. In all the examples, the MCR decreased with increasing HI, indicating that the higher the predicted risk, the lower the number of substantially contributing chemicals. However, the overall assessment relies on the knowledge of the identity of the compounds and their contribution to the overall toxicity.

The lower the MCR, the lower the number of compounds that are significantly contributing to the overall risk. Using the MCR, the "nature of concern" of a mixture and related risk management actions can be decided on. This way, the MCR could equally be considered as a surrogate for the MAF. The MCR does not assume equal toxic share of most contributing substances in a mixture (as in previous section).

Based on the findings in literature and own dataset analysis, one can conclude that, assuming unequal toxic share of most contributing substances in a mixture, a reasonable worst-case MAF factor of 2-3 (90th percentile) would protect the majority of the observed mixtures. This conclusion is made irrespective of whether the observed mixture is at risk or not.

4.3. SIZE OF THE HAZARD INDEX

The 90th and 95th percentile of the Hazard Index was calculated for the EU Waterbase (based on supplemental data from Rhodea-Palomares, in preparation), Rhine, Adour-Garonne (France), the Swedish and the Erft dataset using the chronic HC5 as reference value (see Table 5).

Table 5: High end percentiles of HI (Hazard Index) for observed mixtures (based on a more realistic scenario in which non-detects are excluded). See 2.1.3 for definition of the different mixture group categories.

		For all mixtures (all Groups)	For all mixtures at risk (i.e. Group I and III)	For Group III mixtures	For Group III mixtures + 'risk managed' Group I mixtures
		95%			
Current study	Rhine	1.7	2.9	2.0	2.5
	Adour-Garonne (France)	1.9	6.6	1.7	2.3
	Swedish Pesticides	2.2	14.9	2.1	2.6
	Erft river	13.7	19.9	2.3	4.3
Rodea-Palomares et al. (in preparation)	EU Waterbase	1.3	16.4	2.4	3.3
KEMI report, 2021[a]	Sweden	16.3	NA*	NA*	NA*
	Erft river	42.7	NA*	NA*	NA*
		90%			
Current study	Rhine	1.3	2.4	1.7	2.1
	Adour-Garonne (France)	0.9	4.5	1.6	1.9
	Swedish Pesticides	1.4	7.9	1.8	2.2
	Erft river	9.0	12.2	2.0	3.6
Rodea-Palomares et al. (in preparation)	EU Waterbase	0.7	8.0	2.0	2.6

** NA: Not available: could not be calculated*

[a] Note that in the report of KEMI (2021) PNECs were used as reference values, while the current study considered HC5 values.

It has been argued that a HI>1 together with a MCR close to 1 indicates that a given exposure situation is not relevant from a mixture perspective, but is instead a single substance issue (Price & Han, 2011). Or more general, individual chemicals exceeding safe levels do not need to be addressed specifically in the context of mixtures, as those should be tackled appropriately by the usual single substance legislative measures as postulated by Bopp et al. (2019). The same methodology can be applied on Group III mixtures only since these are the main target mixtures for introducing a MAF in the first place. This analysis was conducted for the EU (Waterbase), Rhine, Adour-Garonne (France), the Swedish and the Erft dataset (see Table 5 column for group III mixtures). MAF (i.e. HI) between 2 and 3 could be derived depending on the data set and the percentage of mixtures covered.

Gustavsson et al. (2017) argues such an argumentation (excluding Group I mixtures) falls too short without taking a closer look at the consequences of single-substance oriented risk managements for the overall toxicity. In order to analyse this issue in more detail, they assumed that single-substance oriented risk mitigation (RM) measures were successfully implemented, leading to a situation in which no individual HQ exceeds a value of 0.95. That is, each component is, after the implementation of risk mitigation measures, assumed to be present at a concentration of a maximum of 95% of its ecotoxicity reference value. Under these circumstances, the analysed aquatic ecosystem would be assessed as having a good chemical status according to the WFD. They then calculated the resulting mixture risk quotients for this situation. The assumed risk mitigations lower risks, but only from a median risk quotient of 2.1 to a median risk quotient of 1.8. Overall, 70% of the sites still have an unacceptably high risk It can be clearly seen that single-substance oriented risk mitigation substantially lowers the overall risks and such measures are thus a critical first step towards a non-toxic environment. But it is also obvious that single substance risk mitigation is unable to ensure that mixture risk is below the critical value of 1. As a consequence of the assumed successful implementation of single-substance oriented risk management measures, the average MCR values of the mixtures increase. This increased evenness is a consequence of introducing a ceiling for the maximum individual HQ at a value 0.95, simply attributing the same risk to all compounds which individually exceeded the ecotox reference value before the assumed risk mitigation.

A number of concerns can be put forward with this rationale:

- Risk management achieving a reduction proportionate to the HQ of the substance is very theoretical. A management practice especially for wide dispersive uses would not modify exposure in one co-exposure sample but positively impact the co-exposure patterns. For a down-the-drain product (e.g. pharmaceuticals), a relevant RMM is a change in WWTP efficiency (upgrade) with a broader impact including the reducing levels of other micropollutants. For example, the introduction of WWTP treatment steps seem to yield a general abatement of the micropollutant load with the % abatement varying by chemistry.
- Also, a more logical next step for group I mixtures is to further refine the mixture analysis e.g. by considering toxic mode of action or other refinements options that are typically available in risk assessment context, such as applying the concentration addition model at the trophic level.

The analysis considering the risk management of individual substance exceedances was, nevertheless, conducted for the Waterbase, Rhine, Adour-Garonne (France), the Swedish and the Erft dataset (see Table 5 last column). It should be noted that these mixtures are somewhat artificial as not every unacceptable risk would be reduced to hazard index of 0.95 but rather lower. HI or MAF between 2 and 4 could be derived depending on the data set and the percentage of mixtures covered.

4.4. EQUAL TOXIC SHARE OF ALL CONTRIBUTING SUBSTANCES

The equal toxic share of all contributing substances method (i.e. the KEMI 2021 algorithm), applied to the datasets of the current study, resulted in MAF values in line with those originally reported in KEMI 2021 (Table 6 and Figure 6). The median MAF value of 6.36 (Rhine dataset) were in line with the reported MAF of 5.16 for the river Erft (KEMI 2021). The Rhine dataset is considered to be the worst-case dataset, closely followed by the Erft dataset, as was the case for other MAF methods as well. The Swedish pesticides database has an intermediate median MAF value (4.11). The Waterbase and the Adour-Garonne dataset had lower MAF values of 1.47 and 1, respectively.
Looking at the MAF95 i.e. the 95th percentile for each dataset, gives an indication which MAF would be protective for 95% of the samples. These values are clearly higher and the Erft (19), Swedish Pesticides (10.8), Rhine (10) and Waterbase (9.16) datasets all have high MAF95 values compared to the Adour-Garonne dataset (3.19).

Finally, the KEMI 2021 algorithm was also applied to the Swedish pesticides monitoring dataset from KEMI 2021 to demonstrate the validity of the implementation of the algorithm. After data selection following the instructions outlined in the KEMI 2021 study (e.g. only data until 2018), the final dataset was slightly larger (1891 vs. 1513 pesticides). However, the analysis showed good concordance with the original results (median MAF of 3.6 and 4.11 for the KEMI 2021 and ARCHE dataset, respectively), indicating the valid implementation of the algorithm (Table 6).

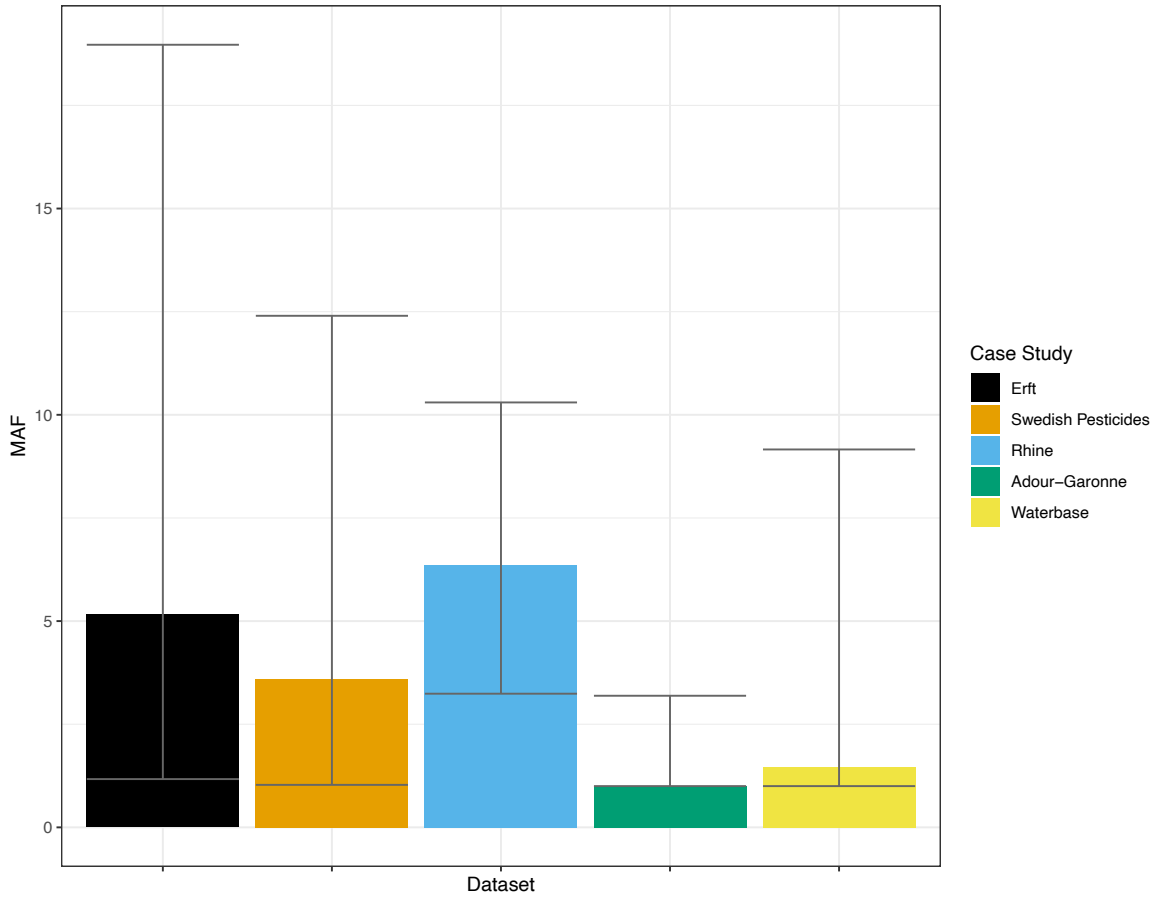

Figure 6: MAF values determined with the equal toxic share of all contributing substances method (cfr KEMI 2021) for different databases. Bars show the median MAF; error bars indicate the 5th and 95th percentile.

Table 6: Results from the mixture assessment factor (MAF) analysis with the KEMI algorithm (2021) for all considered databases. Median values [5th – 95th percentile] are provided. RQSum = sum of risk quotients in the original mixture; Max MAF = maximum MAF for all samples in the database; Number of affected mix components = number of chemicals where the MAF would need to be applied.

Case Study	Number of mixture components	RQSum	MAF	Max MAF	MAF for 90% samples	Number affected mix components	Number of monitored chemicals	Number of samples
Erft river*	25 [4-38]	7.96 [0.05-42.7]	**5.16** [1.17-18.97]	27.5	/	3 [1-16]	/	503
Swedish pesticides (KEMI)*	9 [2-24]	0.88 [0.004-16.3]	**3.6** [1.03-12.4]	31.7	/	3 [1-10]	/	1513
Swedish pesticides (ARCHE)	10 [2-26]	1.08 [0.01-17.01]	**4.11** [1-13.8]	32	10.8	3 [0-11]	109	1891
Rhine (ICPR)	17 [12-22]	3.49 [0.5-7.69]	**6.36** [3.24-10.3]	15.8	9.4	4 [2-8]	57	394
Adour-Garonne (EauFrance)	2 [0-7]	0.01 [0-1.36]	**1** [1-3.19]	5.67	2.45	0 [0-3]	77	3557
Waterbase	4 [0-35]	0.11 [0-7.07]	**1.47** [1-9.16]	31.9	6.4	1 [0-5]	39 [6-121]	12088

* As reported in Table 2 in KEMI 2021.

5. COMPARISON OF THE ROBUSTNESS OF DIFFERENT MAF METHODS

Previous section compared the MAF size for different MAF methods resulting from different databases. Some methods produce similar MAF values while other methods produce quite different MAF values. In order to understand which method is producing realistic values, each MAF method is tested against the main uncertainties of the whole assessment or related to using monitoring data:

- Choice of reference values and built-in conservatism (HC5 versus PNEC)
- Substance sample size including the unknown and non-detected substances potentially contributing to potential mixture risk
- Dealing with mixtures with single substance risks

A robust MAF estimation method should be less sensitive towards these uncertainties.

5.1. IMPACT OF CHOICE OF REFERENCE VALUES

5.1.1. REFERENCE VALUES: HC5 VS PNEC

A comparison of environmental reference values to calculate mixture pressure i.e. chronic HC5 used in Section 4 of this study and PNEC values typically used in other studies (e.g. Gustavsson et al. 2017, Markert et al. 2020, Backhaus 2021) is given in Figure 7 for the Swedish pesticides and the Erft river datasets. With some exceptions, the PNEC is clearly more conservative compared to the HC5. The median difference between the HC5 and the PNEC is 10.7-fold (Swedish pesticides dataset) and 16.6-fold (Erft river dataset). This difference reflects most likely the incorporation of the assessment factor in the derivation of the PNEC (median assessment factor used in the derivation of the PNEC values of the Markert et al. (2020) is 50). HC5 values are preferred above the PNEC in the present study for the estimation of mixture pressure as these are considered to be more relevant to assess risks for actual communities, since HC5s are less influenced by an arbitrary assessment factor (although it should be noted that for some of the SSDs extrapolation factors have been used by Posthuma et al. (2019) to calculate chronic SSD-parameters from acute ecotox data).

Given the large difference between PNEC and HC5 values, the potential influence of the choice of reference value on the derivation of the MAF is further explored in the next sections using the Swedish Pesticides and Erft river datasets as case studies.

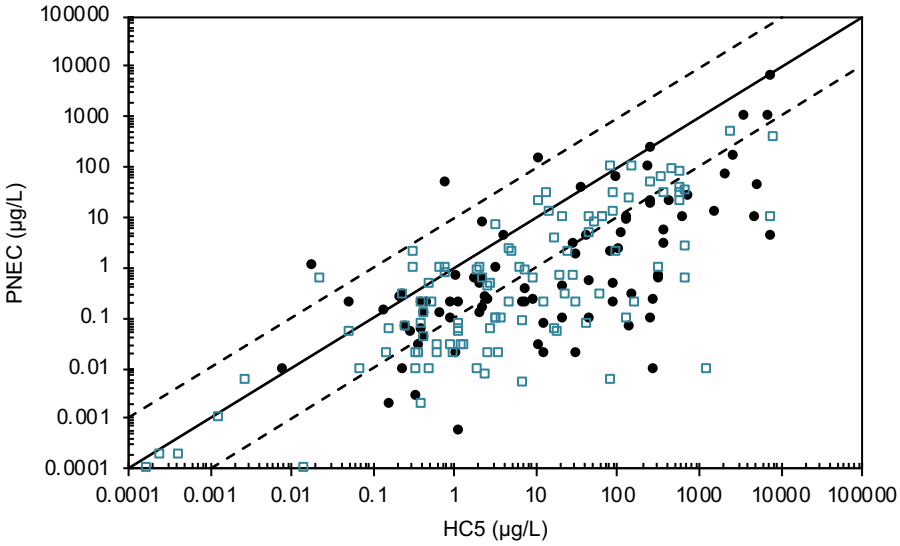

Figure 7 Predicted No-Effect Concentrations (PNEC) as a function of 5% hazardous concentration (HC5) for the Swedish Pesticides (blue open squares) and the Erft river dataset (black circles). Full line represent a PNEC that is equal to the HC5, dashed line represent a 10-fold difference between PNEC and HC5. PNEC were taken from the original studies (Gustavsson et al. 2017, Markert et al. 2020) and HC5 values were derived from the parameters of the chronic SSD reported by Posthuma et al. (2019), see sections 2.5 & 2.6 for more information.

Figure 8 gives an overview of the shift in general toxic pressure in the Swedish database and Erft river dataset when the analysis is based on HC5 vs. PNEC as reference value. Overall, relatively high toxic pressures (expressed as Hazard Index) are observed in both datasets when the PNEC is used as reference value, which can be observed from the percentage of samples that are predicted to be at risk due one or more chemicals to be of concern. The percentage of samples belonging to group II (no risk identified) is notably higher when using the HC5 compared to the PNEC value.

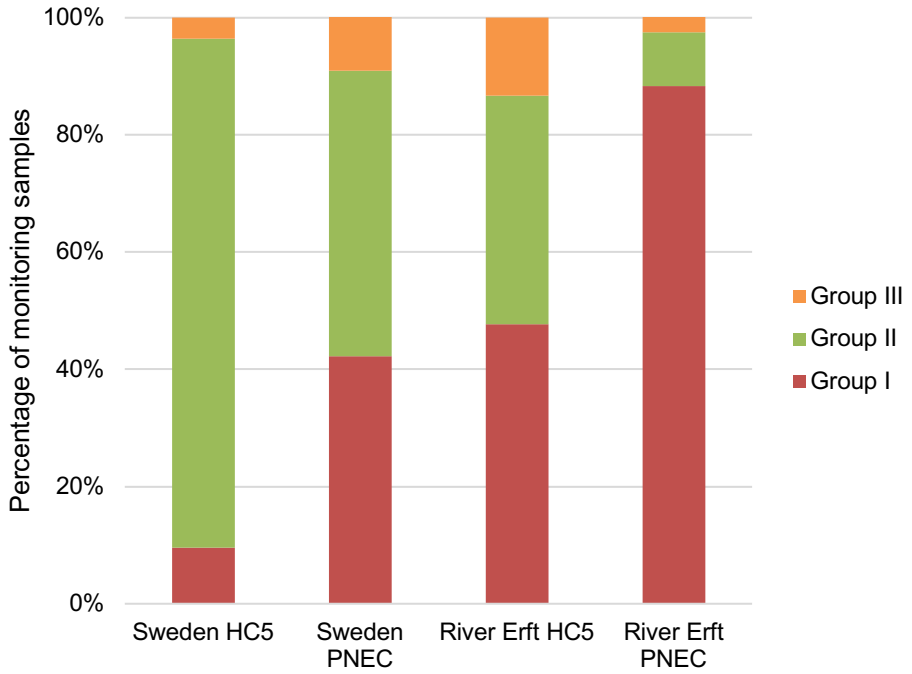

Figure 8 Toxic pressure of mixture samples in the different databases represented by the percentage of monitoring samples belonging to three different mixture groups. Group I are the combined exposures that are a potential concern because one or more individual chemicals are a concern (i.e. at least one of the HQ_i>1). Group II are the combined exposures where there is a low concern for both individual chemicals and for their combined effects: (i.e. HI<1). Group III are the combined exposures where there is a low concern for individual chemicals but there is a potential concern for the combined effects (HQ_i<1, HI>1). Toxic pressure was calculated either using the HC5 or the PNEC as reference value.

5.1.2. IMPACT OF REFERENCE VALUE ON MAF DERIVATION APPROACHES

5.1.2.1. EQUAL TOXIC SHARE OF MOST CONTRIBUTING SUBSTANCES

The contribution of the risk drivers to the Hazard Index calculated with the PNEC as reference value in the Swedish Pesticides and Erft river was calculated and is visualized in Figure 9. The same figures but using the HC5 as reference value are presented in the lower panels of Figure 4. The number of substances contributing to toxicity is relatively unaffected by the choice of reference value. Generally, 3 substances are needed to explain 95% of the hazard index for 50% of the samples, while 4 (Erft river) to 5 substances (Swedish Pesticides) are needed to explain 90% of the hazard index for ca. 90% of the samples. This finding is irrespective of the use of HC5 of PNEC as basis for the Hazard Index calculation.

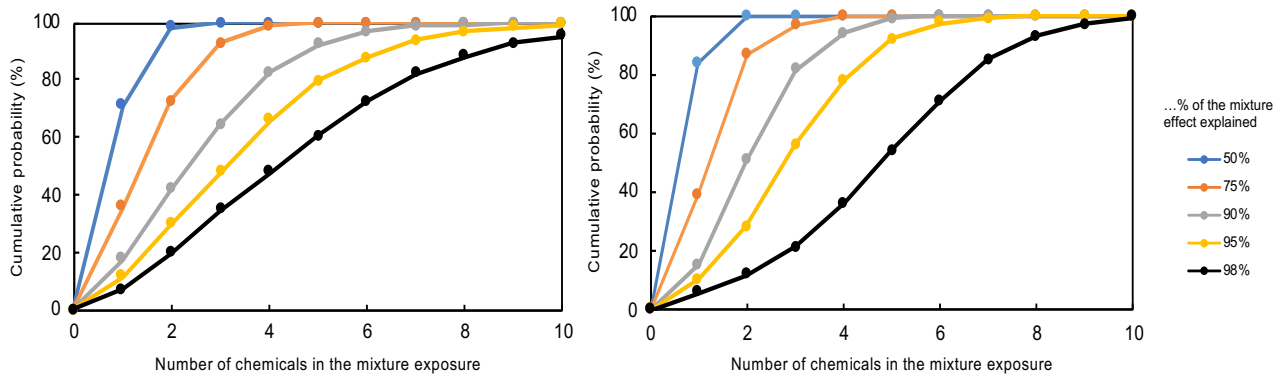

Figure 9: Cumulative probability in function of chemicals contributing for different percentages to the total mixture pressure for the Swedish Pesticides dataset (left panel) and the Erft dataset (right panel). Mixture pressure is expressed as Hazard Index (using PNEC values as reference values (for comparison same figures using HC5 as reference value are given in Figure 4).

Table 7 Number of substances needed to explain a certain percentage of the hazard index for x% of the samples using the HC5 or PNEC as reference value

	Number of substances needed to explain 95% of the hazard index for 50% of the samples		**Number of substances needed to explain 90% of the hazard index for ca. 90% of the samples**	
	HC5	PNEC	HC5	PNEC
Swedish Pesticides	3	3	5	5
Erft river	3	3	4	4

5.1.2.2. MAXIMUM CUMULATIVE RATIO

The Maximum Cumulative Ratio is a measure of the degree of codominance of substances in the mixture, as it is the estimated as the Hazard Index divided by the Hazard Quotient of the highest Hazard Quotient. Values close to 1 indicate dominance of 1 substance in the mixture, while a MCR equal to n indicates that the receptor is exposed to equitoxic doses of all chemicals. The Maximum Cumulative Ratio for the Swedish Pesticides and the Erft river database based on both the HC5 and PNEC as reference value are compared in Table 8. Overall, the MCR is only slightly affected by the choice of the reference value. The reasonable worst-case 90th percentile MCR ranges between 2 and 3 for both datasets, irrespectively of the chosen reference value. The maximum MCR ranges between 4 and 5 for the HC5 and between 3 and 5 for the PNEC. Similar results can be expected for the other databases.

Table 8 MCR or Maximum Cumulative Ratio for representative data sets based on lower tier assessment (based on more realistic scenario in which non-detects are excluded).

	90th percentile MCR		Maximum MCR	
	HC5	PNEC	HC5	PNEC
Swedish Pesticides	2.5	2.7	5.2	5.0
Erft river	2.3	2.0	4.3	3.3

5.1.2.3. SIZE OF THE HAZARD INDEX

The choice of reference value has an important effect on the size of the Hazard Index as represented by the high-end percentiles of HI given in Table 9. When single substance exceedances are taken into account (first and second results columns), the 95% percentile HI calculated using the PNEC as reference value is at least 2-fold higher (but up to 8-fold higher) compared to HI calculated using the HC5 as reference value. On the other hand, the 95% percentile HI of those group of mixtures where HI>1, but no single substance exceedance is observed (i.e. Group III, third results column) is only minorly affected by the choice of reference value. For the Group III + 'risk managed' Group I mixtures, 95% percentile HI are 1.6-fold (Erft river) to 2.3-fold (Swedish Pesticides) higher when the PNEC is used as reference value compared to the HC5 as reference value. Overall, it can be noted that the inclusion of the arbitrary assessment factor inflates the mixture risk predictions when using the hazard index as determinant.

Table 9: Comparison of the influence of the choice of HC5 vs. PNEC for the estimation of cumulative risk of observed mixtures expressed as high-end percentiles of HI (Hazard Index). See 2.1.3 for definition of the different mixture group categories.

		For all mixtures (all Groups)	For all mixtures at risk (i.e. Group I and III)	For Group III mixtures	For Group III mixtures + 'risk managed' Group I mixtures
		95%			
Current study - HC5	Swedish Pesticides	2.2	14.9	2.1	2.6
	Erft river	13.7	19.9	2.3	4.3
Current study- PNEC	Swedish Pesticides	17.0	30.1	2.7	5.9
	Erft river	43.0	46.4	1.5	6.7
Backhaus (2021) PNEC	Sweden	16.3	NA*	NA*	NA*
	Erft river	42.7	NA*	NA*	NA*

5.1.2.4. EQUAL TOXIC SHARE OF ALL CONTRIBUTING SUBSTANCES

The choice of the PNEC or HC5 clearly impacted the calculated MAF value (Table 10). The impact is already apparent when summing the risk quotients (median RQSum of 1.08 and 0.18 for PNEC and HC5, respectively). This was reflected in the calculated MAF values as well: MAF values were lower when using the HC5 compared to the PNEC, both for the median (4.11 and 1.54 with the PNEC and HC5, respectively) and the 95h percentile (13.8 and 6.64 with the PNEC and HC5, respectively).

Table 10: MAF calculations using the KEMI 2021 algorithm for the Swedish Pesticides dataset using either the PNEC or HC5 for RQ calculations.

RQSum = sum of risk quotients in the original mixture; RQSumMan = Sum of risk quotients after setting RQ > 1 to 1; Max MAF = maximum MAF for all samples in the database; MAF 10 coverage = percentage of the samples with RQSum < 1 if MAF = 10; Number of affected mix components = number of chemicals where the MAF would need to be applied.

Case Study	Number of mixture components	RQSum	RQSumMan	MAF	Max MAF	MAF10 coverage	MAF for 95% samples	Number affected mix components
Swedish Pesticides (PNEC)	10 [2-26]	1.08 [0.01-17.01]	1.01 [0.01-4.94]	4.11 [1-13.8]	32	<95%	13.8	3 [0-11]
Swedish Pesticides (HC5)	10 [2-26]	0.18 [0-2.24]	0.18 [0-1.35]	1.54 [1-6.64]	18	>95%	6.64	1 [0-4]

5.2. IMPACT OF SUBSTANCE SAMPLE SIZE INCLUDING NON-DETECTED AND UNKNOWN SUBSTANCES ON MAF

A critical question in the debate on potential risk of unintentional mixtures is how to account for the contribution of unmeasured or unknown substances, as well as undetected substances. This can be assessed by understanding the impact of the substance sample size and how the number of substances contribute to the overall mixture risk. It has been shown that most of the toxicity of even large mixtures is determined by a small number of substances (Pareto principle). This was further investigated for the datasets selected in this study and for the different MAF methods.

5.2.1.2. EQUAL TOXIC SHARE OF MOST CONTRIBUTING SUBSTANCES

The MAF as calculated with the 'equal toxic share of most contributing substances' initially increases for larger samples and levels off starting from intermediate sample sizes. This is especially apparent for the Waterbase (Figure 13) but less clear for the Swedish Pesticides (Figure 12). Therefore, it is expected that the MAF calculated with this method is probably robust against unmeasured and/or undetected chemicals.

5.2.1.3. MAXIMUM CUMULATIVE RATIO

The MCR follows the same pattern as the HI: MCR increases as sample sizes increase but levels off at intermediate sample sizes (see Figure 12 and Figure 13). Reaching a plateau is more pronounced. Consequently, the MAF defined as MCR of a large database is sufficiently robust against the potential impact of unmeasured or undetected chemicals because the MCR focuses on the general observed pattern of substance toxicity contributions in mixtures.

Price & Han (2011) equally demonstrated the MCR is relatively robust against number of detects (and setting non-detects equal to zero) and against number of analytes (and setting non-detects equal to $DL/2^{0,5}$) (see Figure 10).

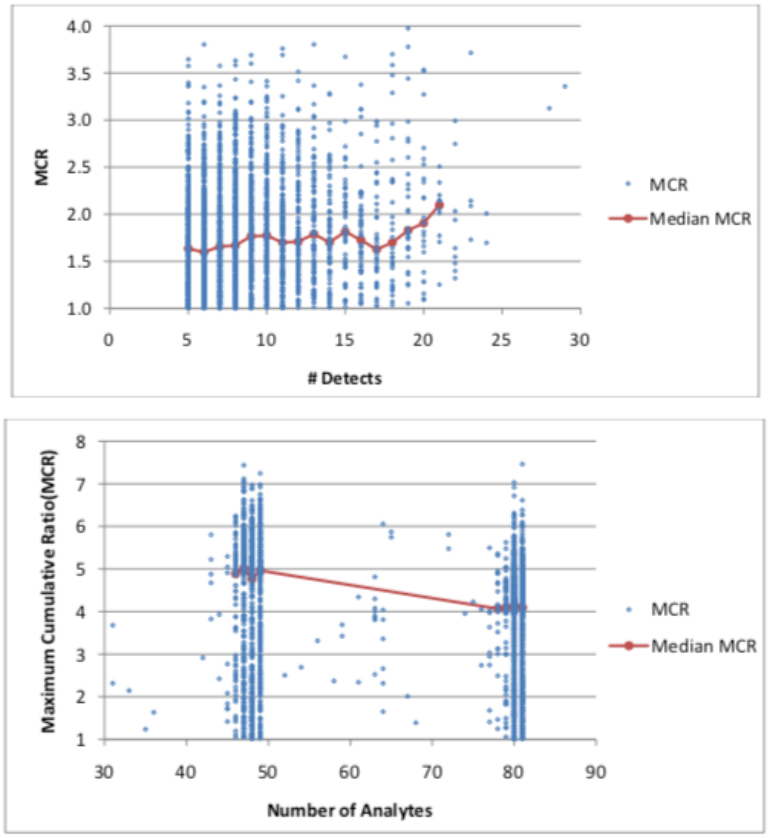

Figure 10: Figures from Price & Han (2011): The relationship between MCR and n for top: scenario 1 (DL = 0) and bottom: scenario 2 (DL = DL//$2^{0,5}$). MCR and n have respectively a weakly positive and even negative correlation.

5.2.1.4. SIZE OF THE HAZARD INDEX

The HI seems to be mostly unaffected by the number of measured substances, even for very large samples (see Figure 12 and Figure 13). Considering all samples in the Swedish Pesticides dataset, the HI starts to increase above 5 substances but remains relatively constant above roughly 25 substances (Figure 12). There are a few outliers above 35 substances, but these only pertain to a few samples. For the Waterbase, samples can be larger (up to 80 substances) but the same trend is confirmed, with an initial increase in the HI but a constant pattern in the HI at higher sample sizes (Figure 13).

This is in contrast with Price & Han (2011) who showed a positive relationship between HI and the number of detects (and equally between the HI and the number of analytes).

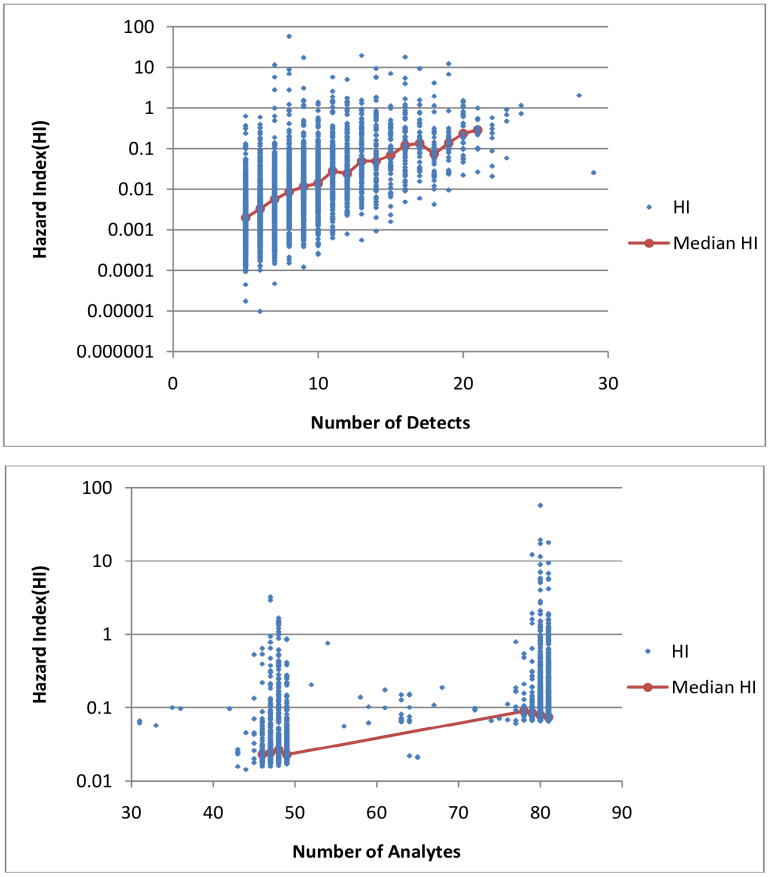

Figure 11: Figures from Price & Han (2011): The relationship between HI and n for top: scenario 1 (DL = 0) and bottom: scenario 2 (DL = DL//$2^{0,5}$). HI and n have respectively a strong positive and weak positive correlation.

5.2.1.5. EQUAL TOXIC SHARE OF ALL CONTRIBUTING SUBSTANCES

The MAF as calculated with the KEMI 2021 algorithm does not reach a plateau as observed for the HI and MCR method: the MAF increases for larger samples, without levelling off as observed for the other two methods (see Figure 12 and Figure 13). The lack of plateau phase for larger samples for the KEMI 2021 MAF does not reflect the observation that a small number of substances drive the mixture risk.

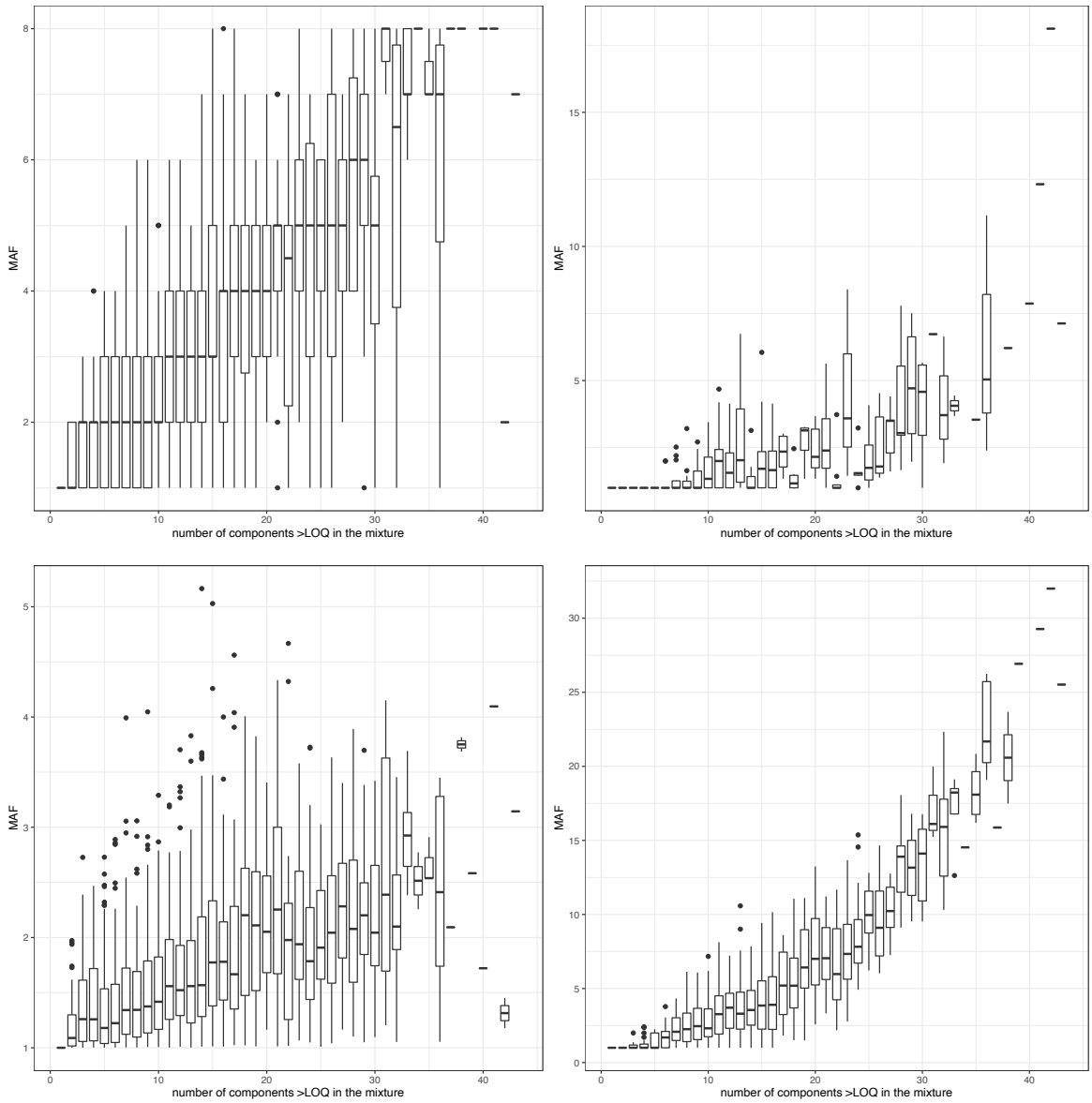

Figure 12: Mixture methodology against the number of components above the limit of quantification (LOQ) in the mixture for the Swedish Pesticides dataset based on four different MAF methodologies: equal toxic share of most contributing substances (upper left, MAF is the number of substances explaining 90% of mixture effect), HI (upper right), MCR (lower left) and equal toxic share of all contributing substances (lower right).

Remark: for easier visualisation, 21 and 6 MAF values higher than 10 were omitted from the plot for HI (left) and MCR (middle) methods, respectively.

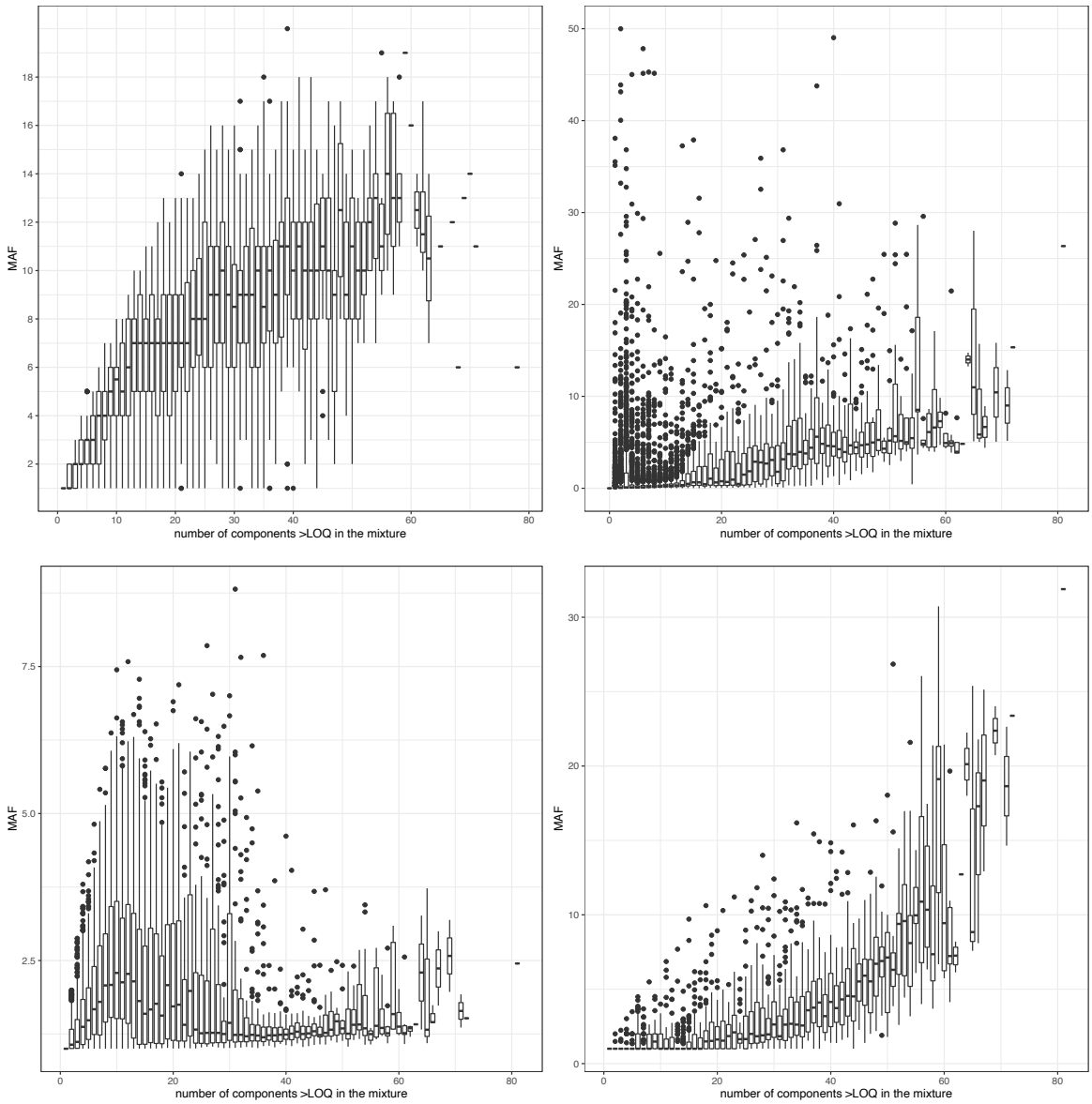

Figure 13: Mixture methodology against the number of components above the limit of quantification (LOQ) in the mixture for the refined Waterbase dataset, based on four different MAF methodologies: equal toxic share of most contributing substances (upper left, MAF represents the number of substances explaining 90% of mixture effect), HI (upper right), MCR (lower left) and equal toxic share of all contributing substances (lower right).

Remark: for easier visualisation, 23 MAF values higher than 50 were omitted from the plot for the HI method (left).

5.3. IMPACT OF SINGLE SUBSTANCES AT RISK

MAF methods differ in how they account for the mixtures where single substances lead to risk. This can be understood from the methodology of each method.

- The equal toxic share for the most contributing substance method looks at the number of substances responsible for the mixture risk, irrespective of the individual risk of the substances.
- The MCR method evaluates the number of substances significantly contributing to the mixture risk. Because this is based on the division between the mixture risk and the highest single substance risk (Eq. 2), the impact of single substance risks in the mixture is limited.
- The size of the hazard index is the sum of the individual hazard quotients of the mixture components. Risks of single substances (HQ > 1) will lead to a higher hazard index and thus have a large impact on the final MAF.
- The equal toxic share of all contributing substances method only considers mixtures at risk. Within the algorithm, the HQ of individual substances at risk is set to 1 before further calculating the MAF. In the calculations, these single substances still have a relatively large impact on the final MAF.

5.4. LEVEL OF PROTECTIVENESS

The level of protectiveness and/or level of conservatism of MAF methods can be evaluated by comparing the MAF values between different levels of protectiveness i.e. the percentage of mixture in the databases that they cover. This is ultimately a policy decision, but typically 90^{th} or 95^{th} percentiles are used for worst-case scenarios. In the current study, we looked at the difference between 90 and 95% of mixtures.

The MCR method was least influenced by this choice, differences between both protection levels were less than 1 across all databases (Table 4). For the equal toxic share for the most contributing substances method, the protection level is accounted for twice and is thus an important factor: in the selection of the % of mixture effect explained and what cumulative percentage of the samples to consider. For the equal toxic share method for all contributing substances method, the impact was notable (MAF difference up to 3 between 90^{th} and 95^{th} percentile). The size of the HI sums the individual hazard quotients and differences between 90^{th} and 95^{th} percentile can thus be considerable, depending on the database (see Table 5).

5.5. SUMMARIZING ROBUSTENESS OF DIFFERENT MAF METHODS

Various methods could be employed to determine the generic MAF. These methods can be based on:

- The number of (equal toxic share) most contributing substances in a mixture, or
- The size of the maximum cumulative ratio, or
- The size of the hazard index (or cumulative risk ratio), or
- Iterative algorithm based on equal toxic share of all contributing substances in carrying capacity

The robustness of all these methods against the following conditions was assessed in previous sections:

- Choice of reference values
- Sample size and the unknown and non-detected substances
- Consideration of mixtures with single substance risks
- Choice of protection level and level of conservatism

A summary of the results in provided in Table 11.

Table 11: Summary of robustness of different MAF estimation methods on monitoring data.

	Equal toxic share of most contributing substances	**Maximum cumulative ratio**	**Size of the hazard index (or mixture risk ratio)**	**Equal toxic share of all contributing substances**
Reference	Van Broekhuizen et al. (2016)	Price & Han (2011)	KEMI (2015)	KEMI (2021)
Logic	Equal toxic share of worst substances in carrying capacity; non-proportional to worst substances	Non-equal toxic share in carrying capacity; proportional to all substances	Non-equal toxic share in carrying capacity, proportional to all substances	Equal toxic share of all substances in carrying capacity; non-proportional to worst substances
Choice of reference values	Robust	Robust	Sensitive	Sensitive
Impact sample size and unknown and non-detected substances	Moderate	Robust	Probably robust	Sensitive
Consideration of mixtures with single substance risks	Robust	Robust	Sensitive	Sensitive
Choice of protection level	Sensitive	Robust	Sensitive	Moderate

and level of conservatism				

It can be concluded that the MAF estimation method **"maximum cumulative ratio"** is **the most robust method** to the choice of reference values, sample size, unknown and non-detected substances consideration of mixtures with single substance risks and choice of protection level. The "equal toxic share of most contributing substances" is slightly less robust due to the higher sensitivity towards sample size and choice of protection level. The "size of the hazard index or mixture risk ratio" and "equal toxic share of all contributing substances" are most sensitive to the defined criteria.

6. CONCLUSIONS

This report does not question or assess whether a MAF is needed, or in which cases it is needed (generic or specific), or whether MAF is the appropriate measure to account for potential mixture risks or whether REACH is the proper regulation to address potential mixture risks. This report assumed that if a generic MAF would be put in place, how to best estimate such generic MAF.

Various methods could be employed to explore a generic MAF. Factors between 2 and more than 50 have been reported. All methods have their own rationale with advantages and disadvantages. The use of spatially broad EU monitoring data is generally preferred as they portray representative, real life occurrence of mixtures. Environmental models are as accurate as a factor 10 to 100 (ratio modelled versus monitored; Van Gils et al., 2020) which is large in the context of MAF setting. Use of environmental monitoring also comes with uncertainties such as non-detected and unknown chemicals in the environment. A proper MAF derivation method should be as robust as possible against these uncertainties. **The method "maximum cumulative ratio" was found to be the most robust against these uncertainties.** The "equal toxic share of most contributing substances" was evaluated as the second most robust method based on the defined criteria. These two methods each rely on a different logic; the equal toxic share of most contributing substances method assumes an equal toxic share of the environmental carrying capacity for most contributing substances and will therefore disproportionally assign higher MAF to the most contributing substances whereas the MCR assumes unequal toxic share of the environmental carrying capacity and proportionally allocates MAF to all substances contributing to the mixture toxicity. **As to which method takes preference is largely a policy decision**. Following questions need to be addressed:

- Should the burden of mixture toxic pressure be distributed to the most contributing substances (equal share of carrying capacity) or to all substances (proportionally, i.e. in which most contributing substances take relative more burden)?
- What protection level is to be achieved, i.e. how many mixtures (sample locations)? In exposure assessment, 90% is often used. In hazard assessment, 95% protection is often used.

The assessment provided above suggests that the two most robust methods (with different policy logic background) result in a reasonable worst-case (ca. 90th percentile) generic MAF factor of 3 (MCR) to 9 (equal toxic share method) to protect the majority of the observed mixtures. These factors refer to all chemicals (i.e. REACH and non-REACH chemicals like plant protection products and pharmaceuticals, as well as historical/legacy substances and sources which are already regulated and for which a MAF will not result in additional benefit). These factors are also **worst-case** because, first, additivity (concentration addition) was assumed as reasonable worst-case interaction. Different toxic modes of action would not necessarily lead to additivity and would result in lower MAF. Secondly, the additivity was assumed at ecosystem level, adding for example effects from fish as most sensitive species for one chemical to effects from invertebrates as most sensitive species for another chemical. A refinement would be to assume additivity at trophic level or community level (e.g. trophic level concentration addition). A REACH specific MAF taking into account independent action between certain groups of chemicals (calculated at trophic level) will be lower. Further differentiation following single substance regulation and/or mode of action is therefore recommended for further research.

7. REFERENCES

Backhaus T, Karlsson M. 2014. Screening level mixture risk assessment of pharmaceuticals in STP effluents. Water Res. 49:157–165.

Bopp SK, Kienzier A, Richarz A-N, van der Linden SC, Paini A, Parissis N, Worth AP. 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward, Critical Reviews in Toxicology, 49:2, 174-189, DOI: 10.1080/10408444.2019.1579169

Devos, Y, Auteri, D, de Seze, G. 2020. EFSA activities on environmental risk assessment (ERA). 75th Advisory Forum Meeting, 1-2- April 2020.

EFSA. 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge of-field surface waters. EFSA Journal 2013;11(7):3290

EFSA. 2015. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 2015;13(2):3996

EFSA. 2016: Scientific Opinion addressing the state of the science on risk assessment of plant protection products for in-soil organisms

EFSA, 2021. EFSA Strategy 2027, Publications Office of the European Union, June 2021. doi:10.2805/886006

Gustavsson, M., Kreuger, J., Bundschuh, M., & Backhaus, T. (2017). Pesticide mixtures in the Swedish streams: environmental risks, contributions of individual compounds and consequences of single-substance oriented risk mitigation. *Science of the Total Environment*, *598*, 973-983.

KEMI. 2015. An additional assessment factor (MAF) – A suitable approach for improving the regulatory risk assessment of chemical mixtures?. PM 5/15.

KEMI. 2021. Improving the regulatory assessment of combination effects: steps towards implementing the mixture assessment factor (MAF) in chemical regulation. PM 8/21. Report prepared by Thomas Backhaus, University of Gothenburg for KEMI, Sweden.

Markert, N., Rhiem, S., Trimborn, M., & Guhl, B. (2020). Mixture toxicity in the Erft River: assessment of ecological risks and toxicity drivers. *Environmental Sciences Europe*, *32*, 1-13.

Posthuma L, De Zwart D, Keijzers R, Postma J. 2016. Watersysteemanalyse met de Ecologische Sleutelfactor Toxiciteit Deel 2. Kalibratie: toxische druk en ecologische effecten op macrofauna. STOWA rapport number: 2016-15 B. Stichting toegepast Onderzoek Waterbeheer, Amersfoort, The Netherlands. ISBN 978.90.5773.727.5

Posthuma L, van Gils J, Zijp MC, van de Meent D, de Zwart D. 2019. Species sensitivity distributions for use in environmental protection, assessment, and management of aquatic ecosystems for 12 386 chemicals. Environmental Toxicology and Chemistry 38, 905–917. https://doi.org/10.1002/etc.4373 (2019).

Price P, Xiangly H, Junghans M, Kunz P, Watss C, Leverett D. 2012. An application of decision tree forassessing effects from exposures to multiple substances to the assessment of human and ecological effects from combined exposures to chemicals observed in surface waters and waste water effluents. Environmental Sciences Europe 24/34.

Price PS, Han X. 2011. Maximum Cumulative Ratio (MCR) as a tool for assessing the value of performing a cumulative risk assessment. International Journal of Environmental Research and Public Health 8:6, 2212-2225. DOI: 10.3390/ijerph8062212

Price PS, Han X. 2011. Maximum Cumulative Ratio (MCR) as a tool for assessing the value of performing a cumulative risk assessment. International Journal of Environmental Research and Public Health 8:6, 2212-2225. DOI: 10.3390/ijerph8062212

Price P, Han X, Junghans M, Kunz P, Watts C, Leverett D. 2012. An application of a decision tree for assessing effects from exposures to multiple substances to the assessment of human and ecological effects from combined exposures to chemicals observed in surface waters and waste water effluents. Environmental Sciences Europe 24(1): 34.

Redman AD, Parkerton TF, Paumen ML, Butler JD, Letinske DJ, den Haan K. 2017. A re-evaluation of PETROTOX for predicting acute and chronic toxicity of petroleum substances. Environmental Toxicology and Chemistry 36:8, 2245-2252.

Rodea-Palomares I, Weyers, A. 2021. From Exposure to Risk: What Can We Learn From Large Scale Chemical Monitoring and Chemical Mixtures ERA? Platform presentation at SETAC Europe 31th Annual Meeting.

Rodea-Palomares et al. *in prep*.

Salminen R. (Chief-editor), Batista MJ, Bidovec M, Demetriades A, De Vivo B, De Vos W, Duris M, Gilucis A, Gregorauskiene V, Halamic J, Heitzmann P, Lima A, Jordan G, Klaver G, Klein P, Lis J, Locutura J, Marsina K, Mazreku A, O'Connor PJ, Olsson S.., Ottesen RT, Petersell V, Plant J.A, Reeder S, Salpeteur I, Sandstrom H, Siewers U, Steenfelt A, Tarvainen T. 2005. Geochemical Atlas of Europe. Part 1: Background Information, Methodology and Maps. Geological Survey of Finland, Espoo, Finland. Available from: http://weppi.gtk.fi/publ/foregsatlas/

Tarazona J. 2016: Towards the integration of ecological and landscape diversity in prospective ERA. EFSAScientific Conference, November 2016.

Vallotton N., Price, P. S., 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. Environ. Sci. Technol. 50, 5286−5293. doi:10.1021/acs.est.5b06267

Van Broekhuizen, FA, Posthuma L, Traas TP. 2016. Addressing combined effects of chemicals in environmental safety assessment under REACH - A thought starter. RIVM Letter report 2016-0162. https://www.rivm.nl/bibliotheek/rapporten/2016-0162.pdf.

Van Gils, J., Posthuma, L., Cousins, I., Brack, W., Altenburger, R., Baveco, H., Focks, A., Greskowiak, J., Kuehne, R., Kutsarova, S., Lindim, C., Markus, A., van de Meent, D., Munthe, J., Schueder, R., Schuurmann, G., Slobodnik, J., de Zwart, D., & van Wezel, A. (2020). Computational material flow analysis for thousands of chemicals of emerging concern in European waters. Journal of Hazardous Materials, 397, 122655. https://doi.org/10.1016/j.jhazm at.2020.122655

Verro, R., Finizio, A., Otto, S. & Vighi, M., 2009. Predicting pesticide environmental risk in intensive agricultural areas. II: Screening level risk assessment of complex mixtures in surface waters. Environmental science & technology, 43(2), pp.530–7. doi:10.1021/es801858h