STE methodology for monitoring-based prioritisation of chemical substances in support to EU water policy

The Technical Report is the JRC proposal to the SG-R group about an update of STE method in compliance with the lower measurement data quality

Raquel N. Carvalho, Dimitar Marinov, Lidia Ceriani, Robert Loos and Teresa Lettieri

September, 2015

Table of Contents

Introduction

Directive 2000/60/EC (WFD) establishing a framework for European Community action in the field of water policy lays down a strategy for the prevention and control of chemical pollution of surface and coastal zone waters. The second revision of the Priority Substances (PS) and Priority Hazard substances (PHS) list was initiated in 2014. The ongoing prioritisation approach follows the previous exercise (2007-2011) in distinguishing monitoring data rich substances going through the monitoring-based exercise from the monitoring data poor substances going through the modelling-based one. However, a new monitoring-based prioritisation methodology (STE) was developed, based on the scheme developed by von der Ohe et al. (2011), which exploits the inherent variability in the monitoring dataset, and evaluates the Spatial, Temporal and Extent of PNEC exceedances, in order to rank and classify the substances for the risk they pose to the European surface waters.

STE considers the complete variety of environmental matrixes (water, sediments, and biota) and water bodies (surface, ground or drinking water). In particular for surface water it differentiates inland water (rivers, lakes and artificial channels and reservoirs) from transitional and coastal water, and within those water types it distinguishes the different fractions (whole water, dissolved phase, SPM).

Similarly to the classical chemical risk assessments (James et al., 2009), STE estimates Risk Quotients (RQ) as a ratio of measured concentrations in a given environmental matrix to the toxicological threshold of concern (EQS or PNEC). Then, STE calculates Spatial, Temporal and Extent factors of exceedances per substance for measurements in EU and EFTA countries and combines these factors in a single and representative ranking score. Next, STE finds the scores for chemicals by summing their Spatial, Temporal and Extent exceedance factors:

$$Score = F_{spatial} + F_{temporal} + F_{extent}$$

The range of Scores is between 0 and 3 (since the individual factors vary from zero to one), where 0 means no concern, and 3 indicates a very high concern.

The factor of spatial frequency of exceedance for a given substance is quantified as a product of two proportions – one related to sites and a new one that specifies a correction by the number of affected countries:

$$F_{spatial} = \frac{EXC_{\max,site}}{\#Site} \cdot \frac{EXC_{\max,country}}{\#Country}$$

where $EXC_{max,site}$ accounts the number of sites in which the max concentration exceeded the PNEC; #Site is the number of sampling stations in which the substance was measured; $EXC_{max,country}$ gives the number of countries in which the max concentration exceeded the PNEC; and #Country is the number of countries in which the substance was measured.

The newly developed factor of the temporal frequency of exceedances is formulated as:

$$F_{temporal} = \frac{\sum_{SiteECX} \frac{EXC_{sample}}{\# Analysis}}{\# SiteECX}$$

where EXC_{sample} represents the number of samples in one site which exceeded the PNEC, *#Analysis* shows the total number of analyses in one site, and *#SiteECX* gives the total number of sampling stations in which the measured concentrations exceeded the PNEC at least once. If *#SiteECX* is zero (no stations with exceedances) then $F_{temporal}$ is automatically set to zero.

The factor of extents of exceedances is calculated as:

$$EXC_{extent} = \begin{cases} 95^{th}\ percentile\ (RQ_{max,site}) & \# sites \geq 20 \\ MAX\ (RQ_{max,site}) & \# sites < 20 \end{cases}$$

where $RQ_{max,site}$ is the max risk quotient and F_{extent} is quantified by a normalisation of values of EXC_{extent} according to a new and different lookup table as shown below

EXCextent	<1	1.-2.	2.-5.	5.-10.	10.-20.	20.-50.	50.-100.	100.-500.	500.-1000.	>1000
Fextent	0	0.04	0.07	0.11	0.18	0.28	0.41	0.56	0.75	1

Table 1. New look-up table for quantification of the Extent factor.

Trial tests of STE were performed using PS measurements from SoE and MSDAT datasets and PNEC values equal to the EQS (Directive 2013/39/EU). SoE (State of the Environment – WISE) is managed by the EEA. MSDAT data were directly submitted to the JRC by MS following a request of DG ENV to the EU Water Directors (on 21th March 2014). However, the trial STE exercise evidenced that the quality of measurement data is not satisfactory (doubts for errors, wrong units, bad recording, typos, secondary contamination of samples, sampling or analytical problems, etc.).

The purpose of this work is to discuss an update of STE method in compliance with the found lower data quality. The proposed modification is based on a detailed investigation of the robustness of the STE factors, including their inter-correlations, and a sensitive analysis of STE outcome depending from the number of sampling stations and records and the PNEC uncertainties.

Investigation of robustness of STE

Fspatial

The additional country term in $F_{spatial}$, proposed by JRC, grabs the impact of the spatial propagation of toxic chemicals when they are observed in several countries across Europe. This term is making a correction by the number of affected countries, and plays a role by weighting the score quantified by sites. Indeed, the order of magnitude of the country term and its impact on $F_{spatial}$ can be seen in Figure 1. The figure also evidences that $F_{spatial}$ follows entirely the variability and magnitude of the site term, while the country term imposes on it an average reduction by 0.58 (ranging 0.04 - 1). The bigger discrepancies were observed for the small spatial factors (when $F_{spatial} < 0.1$) whereas for higher ones ($F_{spatial} \geq 0.1$) the reduction effect is only 18%.

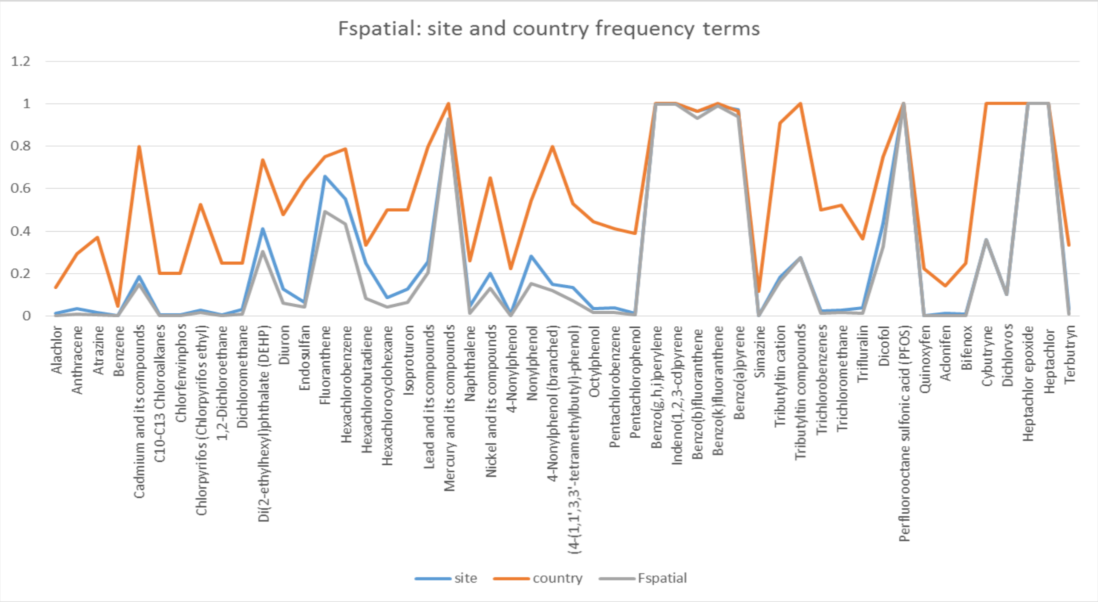


Figure 1. Magnitude, variability and impact of the country and site terms on Fspatial

In addition, since some concerns were expressed that the introduction of the country term would modify substantially the $F_{spatial}$, it was investigated the impact of taking a square root of the product of the site and country terms. The comparison of $F_{spatial}$ calculated only by the site term (indicated as “no country term”) with the “square root” and “with country term” (as originally proposed by JRC) is shown in Figure 2 and Table 2. This evidenced that the square root option is not acceptable as it increases unrealistically $F_{spatial}$ instead of having the expected opposite effect.

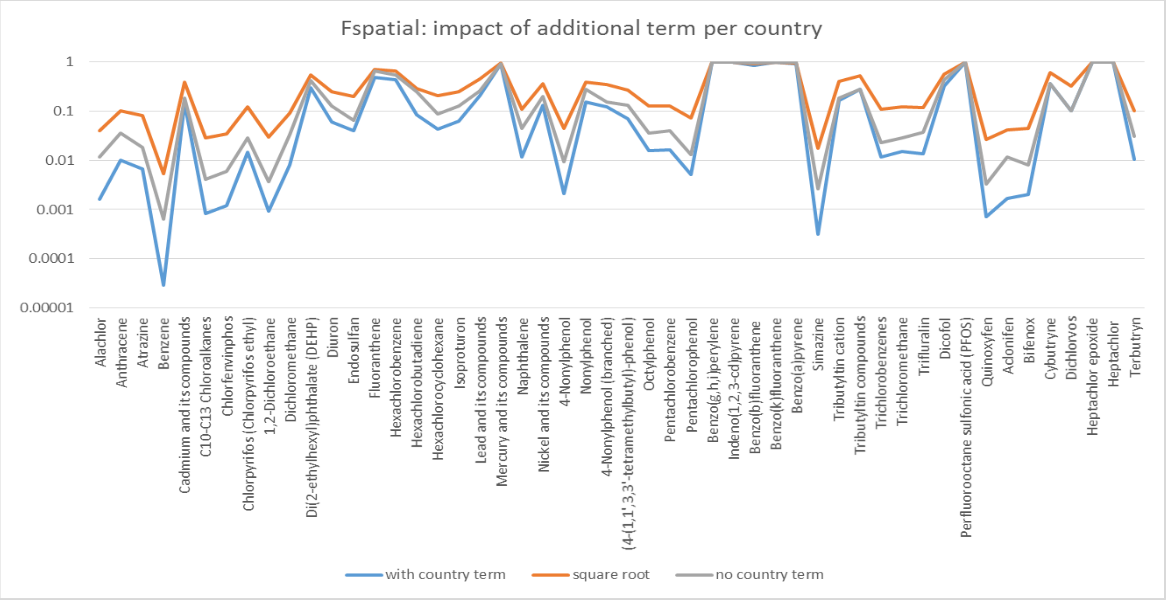


Figure 2. Comparison of $F_{spatial}$ calculated only by the site term (indicated as "no country term") with the "square root" and "with country term".

	min	**average**	**max**
Ratio of (with country term)/(no country term)			
Fspatial (all values)	0.04	0.58	1
Fspatial > 0.1	0.33	0.82	1
Ratio of (squared root)/(no country term)			
Fspatial (all values)	0.96	3.01	8.43
Fspatial > 0.1	0.96	1.51	3.14

Table 2. Comparison of $F_{spatial}$ calculated only by the site term (indicated as "no country term") with the "square root" and "with country term"

JRC investigated the impact of 99th and 95th percentiles as an alternative to the max concentrations in STE when calculating exceedences per site or country in $F_{spatial}$. The obtained results showed a reduction of $F_{spatial}$ in comparison with max option on average by 43% and 61% for 99th and 95th percentiles, respectively. This is a considerable impact, thus it was concluded that the use of max concentrations is a reasonable choice under the condition the measurement data are exempted of outliers and errors. However, when the quality of measurement data is not satisfactory (doubts for errors, wrong units, bad recording, typos, secondary contamination of samples, sampling or analytical problems, etc.) a better choice would be for the 95th percentile of measured concentrations at a given station.

Ftemporal

For a proper calculation of the temporal contribution a key factor for each sampling station is the overall number of samples per substance since, if it is very low, then the ratio will approach to a value of 1 (one) which could not reflect neither the real scenario nor a temporal trend . For example it was found that on average, the PS, observed only once for the entire study period (2006-2014) was at 9.1% of the stations. For this reason JRC introduced a two-step procedure for calculation of $F_{temporal}$. Accordingly, the first approximation is calculated considering all observation sites. Next, the second estimate is found by excluding stations with less than 2 measurements when PNEC exceeded. Then, the final value of $F_{temporal}$ equals to the first approximation if the deviation between both estimates is below 10% and vice versa in the opposite case.

Fist estimation – $f1$ Second estimation - $f2$	calculated considering all observation sites excluding stations with less than 2 samples when PNEC is exceeded
Final $F_{temporal}$	If $\lvert (f1-f2)/f1 \rvert <= 0.1$ then $F_{temporal} = f1$ If $\lvert (f1-f2)/f1 \rvert > 0.1$ then $F_{temporal} = f2$

Table 3. A two-step procedure for calculation of $F_{temporal}$

Fextent

For a better quantification of the F_{extent} factor a more detailed and flexible lookup table is needed that comprises additional levels for EXC_{extent}. The comparison of F_{extent}, calculated by the classical (von der Ohe et al. (2011)) and our new approach, is presented in Figure 3. It shows for single substances a different variability - from reduction by half up to two-times growth, but apparently the F_{extent} increases more gradually and behaves more smoothly.

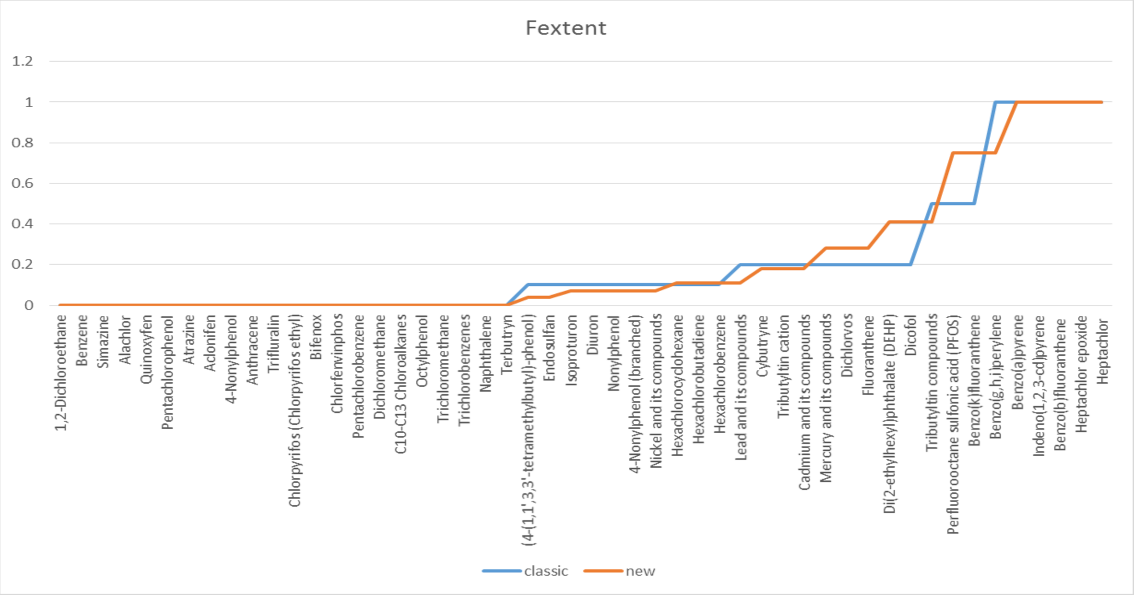


Figure 3. Comparison of *Fextent* calculated by the classical (von der Ohe et al. (2011)) and the new estimate of the STE approach.

Next, the dependence of EXC_{extent} and F_{extent} from the percentiles (90 or 99 compared to 95) and Risk Quotients (percentiles 95 or 99 compared to max) has been investigated (see the formula of the Extent factor).

Under the different percentiles, the EXC_{extent} on average showed (see Table 4) a reduction by 50% for 90-percentile when comparing to 95-percentile, and a big expansion of 3-4 orders of magnitude for 99-percentile. This explains the observed alike numbers of F_{extent} for the 90- and 95-percentiles (on average only 18% deviation) and the unrealistic jump of F_{extent} for several PS from zero (or values close to zero) to one in case of 99-percentile (on average 104% deviation). Thus, the correctness of the choice of 95-percentile in the formula for the extent of exceedances has been confirmed.

	EXC_{extent}	F_{extent}
90-percentile	50% reduction	alike numbers (on average 18% deviation)
99-percentile	expansion of 3-4 orders	big differences (on average 104% deviation)

Table 4. Impact of using 90- and 99-percentile instead of 95-percentile on Extent factor

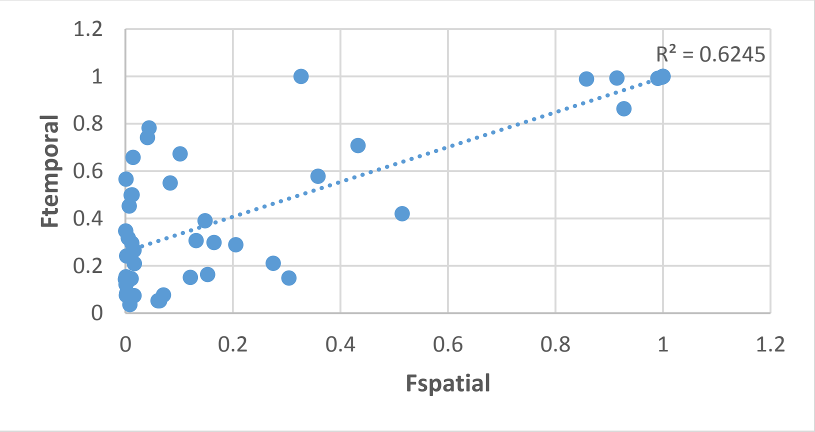
The investigation of the impact of the Risk Quotients (see Table 5) evidenced that the EXC_{extent} values have been reduced by 7% and 30%, when using 99- and 95-percentiles, respectively. Nonetheless, the three scenarios for the Risk Quotients showed similar results for F_{extent} - for example the deviation is 4.9% for 99-percentile and 15.6% for 95-percentile in comparison to the max option. In conclusion, the selection of max Risk Quotients in the formula of the extent factor is a reasonable decision. However, likewise to the other factors, when the quality of measurements is not satisfactory JRC proposes to substitute the max value with the 95-percentile.

	EXC_{extent}	F_{extent}
99-percentile	7% reduction	on average 4.9% deviation
95-percentile	30% reduction	on average 15.6% deviation

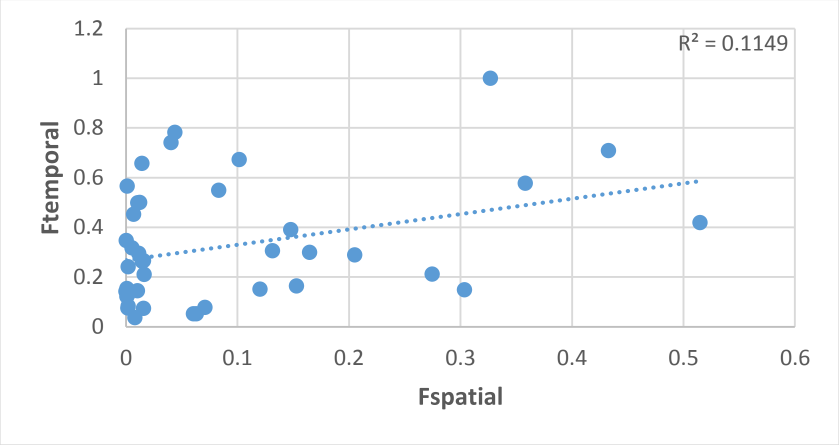
Table 5. Impact of using 99- and 95-percentile instead of max value in RQ on Extent factor.

Correlation between factors

The linear correlation between Spatial and Temporal factors is shown in Figure 4(a) as an example. We noticed that a possible reason for the observed relatively higher correlations between STE factors is the presence of several dots situated in the zone close to one - indeed, the correlation $F_{spatial} \sim Ft_{emporal}$ becomes very low without the outliers (see Figure 4(b)). These dots correspond to the substances with the STE factors ~ 1 which apparently leads to higher correlations. This allows concluding that the STE factors are satisfactory independent from practical point of view.



(a)



(b)

Figure 4. Correlation between Spatial and Temporal factors for PS (a) – all values and (b) – without outliers.

Sensitive analysis of STE

Number of samples and sites

The STE scores for PS found for both scenarios - only quantified samples (concentration is higher than LOD or LOQ) (Scenario1) and all records (Scenario2) are shown in Figure 5 together with the percentage of non-quantified measurements as a part from the total number of samples converted as numbers between 0-1 (if the value is expressed as percentage the figure is not well readable).

As expected when the numbers of samples for both scenarios are almost the same (the number of less-than measurements is low (or zero) the STE scores of both scenarios are identical (or equal). However, if the less-than measurements dominate (for example they are more than 60% from the total number of the records) the Scenario 1 always generates a higher scores - on average 3.1 times (within a range 1-15). This effect is caused by the reduced number of samples or stations in which a given substance was measured above the LOQ which leads to an increase of the Temporal and Spatial factors. In addition, the higher scores for some PS under the Scenario 1 were induced by the expanded Extent factors as a results of the higher 95-percentiles calculated for a shorter time series of data and the following rise of Risk Quotients.

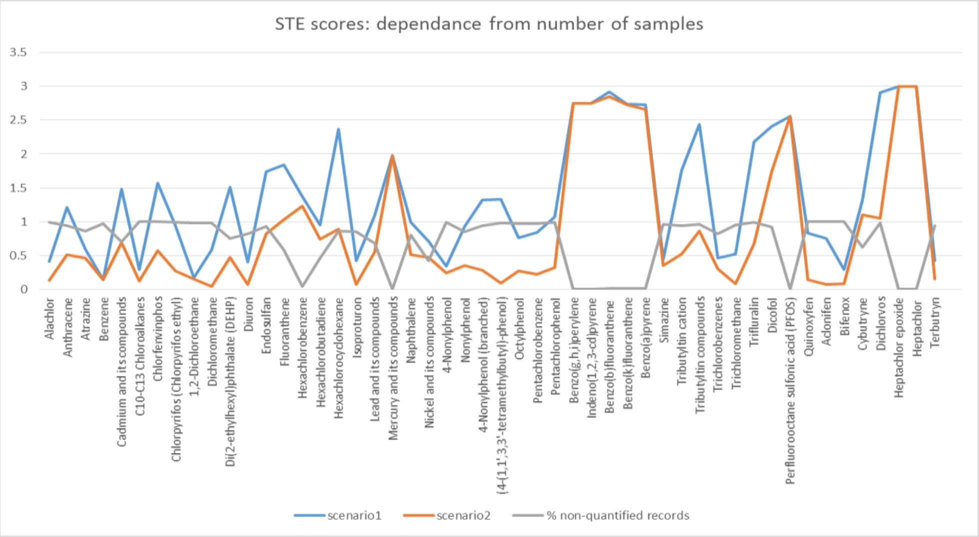


Figure 5. STE scores for PS found for two scenarios - only quantified (concentration is higher than LOD or LOQ) (Scenario 1) and all records (Scenario 2) together with the percentage of non-quantified measurements as a part from the total number of samples given as numbers between 0-1 for convenience.

The results confirm the sensitivity of the STE method to the number samples and sites where substances are measured, thus it is important to ensure data with a sufficient number of measurement stations and records per substance and to dismiss situations when a given substance is measured occasionally or just once at individual sites. However, the existing datasets contain satisfactory number of records, thus the JRC is not considering the introduction of additional eliminating criteria to the data quality.

PNEC

It is well known that the PNEC-values are a crucial element in the chemical risk assessment (James et al., 2009; Dulio and von der Ohe, 2013). In order to evaluate the influence of PNEC on the overall STE scores, the method has been tested with two artificial PNEC sets – ten times lower and higher, respectively. Apparently, this is a pure mathematical drill aiming to evidence the STE model response to uncertainties in PNECs. The obtained results are shown in Figure 6.

Expectedly for most of the PS the reduction of PNECs leads an increase of STE scores and *vice versa*. For instance, the ten times lowering of PNECs displayed on average 5.72 times increase (ranging from 1.03 up to 33.05) while the opposite ten times PNECs enlargement causing on average 0.62 reduction (with a range 0.001-0.955) of the scores.

These values confirm the sensitivity of the STE method to the PNEC choice and the uncertainties in the PNEC values.

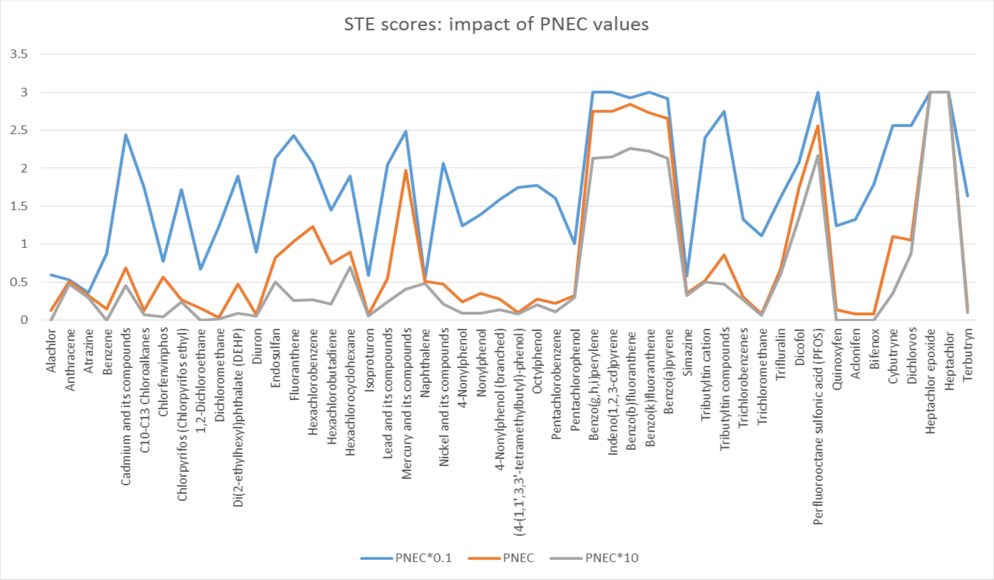


Figure 6. Comparison of STE scores obtained using PNEC values corresponding to the EQS Directive (baseline Scenario2) and two artificial PNEC sets – ten times lower and higher, respectively. The purpose of this pure mathematical drill is to evidence the model response to uncertainties in PNECs.

Conclusions

STE is a new monitoring-based prioritisation methodology, which exploits the inherent variability in the monitoring data, and evaluates the Spatial, Temporal and Extent of PNEC exceedances, in order to rank and classify the substances for the risk they pose to the European surface waters. The STE is built on a scheme developed by von der Ohe et al. (2011) but it ameliorates the Spatial and Extent factors and extends further the original method by inclusion of an additional temporal factor.

The investigation of STE robustness showed that:

- the usage of max concentrations is a reasonable choice when measurement data are exempted of outliers and errors. However, if the quality of measurement data is not satisfactory (doubts for errors, wrong units, bad recording, typos, secondary contamination of samples, sampling or analytical problems, etc.) JRC proposes the max to be substituted with 95th-percentile of measured concentrations at a given station
- the STE factors are not correlated and are satisfactory independent from practical point of view and there is not causational relation among them.

The sensitive analysis of STE evidenced that:

- the method is sensible to the number samples and sites where substances are measured, thus it is important to ensure data with a sufficient number of measurement stations and records per substance and to dismiss situations when a given substance is measured occasionally or just once at individual sites. Therefore, JRC proposes the introduction of an two-step procedure for calculation of Temporal factor of STE based on at least 2 measurements at each station per substance.
- the method depends on the choice of and the uncertainties in the PNEC values and showed their importance as input parameter in the prioritisation process.

Finally, STE is a new monitoring-based prioritisation methodology, which exploits the inherent variability in the monitoring data, and evaluates the Spatial, Temporal and Extent of PNEC exceedances, in order to rank and classify the substances for the risk they pose to the European waters. STE is built on a scheme developed by Peter von der Ohe et al. (2011) but JRC has ameliorated the Spatial and Extent factors and improved further the original method by inclusion of an additional Temporal factor. STE methodology delivers a new and improved tool and scoring system for monitoring-based prioritization of chemicals at continental scale in support to EU environmental legislation.

References

Dulio V., P.C. von der Ohe (editors), 2013. NORMAN Prioritisation framework for emerging substances, NORMAN Association N° W604002510, p.61, ISBN : 978-2-9545254-0-2

EC, 2013. Directive 2013/39/EU of the European Parliament and of the Council amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy, Official Journal of the European Communities, L 226, 24.8.2013, p.17

James A, Bonnomet V, Morin A, Fribourg-Blanc B. Implementation of requirements on priority substances within the context of the Water Framework Directive. Prioritization process: Monitoring-based ranking, 2009. p. 58.

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