NORMAN Working Group on Prioritisation of Emerging Substances

# Prioritisation framework for emerging substances: critical review

## Discussion paper

## WG-1 Meeting, Leipzig, 29 Nov 2017

Prepared by Valeria Dulio, Florian Vergnaud, Nikiforos Alygizakis, Juliane Hollender, Emma Schymanski, Peter von der Ohe

# Table of contents

Table of contents ..... 2

1 Feedback from implementation of the current Prioritisation framework and objectives of the review ..... 4

2 NORMAN initiatives already in place in support of prioritisation ..... 5

3 List of substances: from hundreds, to tens of thousands candidate substances, how to organise the categorisation / prioritisation process for a much larger list of compounds ..... 8

4 How to deal with multiple commercial products/CASRN associated with analysed compounds ..... 9

5 Monitoring data for prioritisation: how to identify and treat outliers ..... 11

6 Improving categorisation of the list of candidate substances ..... 13

6.1 Reviewing the criteria to define “sufficiently” monitored / “sufficiently” quantified substances ..... 14

6.1.1 Application at the EU level ..... 14

6.1.2 Application at the national level ..... 15

6.2 Refining the criteria for categorisation of substances with sufficient monitoring data 16

6.2.1 Revising criteria for categorisation of substances with sufficient monitoring data but low frequency of quantification ..... 16

6.2.2 Revising the distribution of well-monitored substances between Categories 1 and 6 17

6.2.3 For better accounting of the relevance of Cat 1 substances: European or national relevance ..... 20

6.2.4 For better accounting of PNEC uncertainties: refining criteria for Category 3 ..21

6.3 New criteria for categorisation of insufficiently monitored compounds, taking into account NTS data and other information sources ..... 22

7 Improving the ranking algorithm ..... 28

7.1 Exposure score ..... 28

7.1.1 Revising the “Exposure score” (monitoring data) for Categories 1,3,6 ..... 28

7.1.2 Revising the “Exposure” score for categories 2,4 and 5 ..... 31

7.2 Hazard Score ..... 35

7.2.1 A multi-criteria approach for calculation of the hazard score (based on PNEC and CMR) 35

7.3 How to justify scores associated with default values? ..... 37

8 New indicators to improve prioritisation of compounds contributing to mixture effects (potential frequent contributors) ..... 37

References ..... 39

# 1 Feedback from implementation of the current Prioritisation framework and objectives of the review

Prioritisation of chemical contaminants remains a task of primary importance for environmental managers and for the scientific community as regards the definition of priority actions for pollution prevention & control and for the allocation of resources to address current knowledge gaps. It is widely recognised that the lack of data is the primary cause of the lack of regulation of contaminants of emerging concern, as a result of the vicious circle where: “no monitoring means no data, and no data means no regulations”.

The NORMAN prioritisation framework [1] provides a powerful integrated strategy to take gaps into account in the prioritisation of chemical contaminants.

The NORMAN prioritisation concept combines the traditional risk-based ranking process with the *preliminary* application of a decision tree, which allows the allocation of substances into six action categories, based on the knowledge gaps and actions needed to fill them, e.g. development of more powerful analytical methods, launch of monitoring campaigns, performing additional ecotoxicity tests. The ranking within each category is then evaluated by occurrence, hazard and risk criteria. This is a transparent and rational approach to deal with the knowledge gaps which still prevent, for most emerging substances, proper risk assessment and risk ranking.

The NORMAN scheme has been in place since 2013. It has been adopted by WG-1 to provide recommendations to the European Commission for the selection of the substances for the first EU Watch List and the review of the list of Priority Substances under the Water Framework Directive) [2]. It has been implemented in France ([3], [4]) and within the Danube River Basin (14 countries) [ref.] and is currently being studied for possible application in the Netherlands.

New challenges: why and where do we need to improve the current priorisation methodology?

We are increasingly aware that we need to deal with several tens of thousands of chemicals and that we need to be able to take into account the combined effects of this multitude of chemicals as they enter the environment and the food chain, even though each chemical used in a minute quantity may be considered harmless.

The scheme is based on a top-down approach which has so far been applied to the list of “NORMAN emerging substances” – about 800 substances – using data from target monitoring together with substance hazard and physico-chemical properties as the main information source. However, it is today recognised that this list is relatively limited when the considering the multitude of chemical compounds present in the environment. We need to evolve towards a system able to deal with several thousands of compounds and for that, new steps must be introduced in the categorisation / priorisation algorithm to allow better integration of the results from new monitoring-based approaches, such as suspect and non-target screening (NTS) as well as effect-based methods (EBM).

In this context the main aim of WG-1 for the coming years is:
1) to make the NORMAN prioritisation scheme more effective in terms of retrieval and use of supporting data for a much larger list of chemicals (several thousands) for prioritisation of substances and
2) to prepare the evolution towards an integrated scheme combining a large variety of “substance-based” information sources and “monitoring-based” approaches for a more

"robust" identification of priority substances at larger scale and a better understanding of how to reduce current knowledge gaps on emerging contaminants.

In the light of the experience acquired with the implementation of the NORMAN categorisation / prioritisation scheme over the past years, it is proposed to carry out a critical analysis of:

- the strong and weak points of the categorisation algorithm (e.g. cut-off values, criteria for allocation of the substances to the different categories),
- the indicators used for ranking of the substances within each category,
- the procedure for the regular update of the list of highest priority substances in each category, taking into account input from monitoring data (including NTS data) and new effect data,
- the required modifications of the current scheme for better consideration of mixture effects in the prioritisation process.

The final aim is also to decide within the NORMAN WG-1 what type of improvement is needed in connection with the programming of the prioritisation algorithm in the EMPODAT database and automated retrieval of data from other information sources. This activity will involve one WG meeting in 2017 and further meetings in 2018 (in addition to exchange by e-mail) with the aim of publishing an improved version of the NORMAN Prioritisation scheme at the end of the review process (expected end 2018).

**In the following chapters, specific aspects of the NORMAN categorisation / prioritisation scheme are critically analysed. In each chapter, the current NORMAN methodology is first set out, and the critical issues and the proposed changes are then presented to prepare the discussion at the next WG-1 meeting.**

# 2 NORMAN initiatives already in place in support of prioritisation

*SusDat*

The aim of the NORMAN Suspect list exchange initiative is to share information on substances among different reference laboratories, with a view to their use for retrospective screening studies.

In practice, various lists were contributed by different NORMAN partners and NORMAN-connected initiatives (e.g. STOFF-IDENT, list of antibiotics from ANSWER ITN project) and new lists are regularly provided by various contributors. For each list specific information is provided about the origin and the contributor.

The lists are pulled together and curated (removal of duplicates, removal of salts, etc.) in order to end up with one merged list of unique environmentally-relevant substances, which forms today the "NORMAN SusDat" database [5], available on the NORMAN website at http://www.norman-network.net/datatable/. The "NORMAN-SusDat" database contains information for each compound (name of the substance, molecular formula, structural identifiers, connection to other compound databases, predicted toxicological properties, physico-chemical properties and additional mass spectral information useful for the identification of the compounds).

*DSFP "Digital Sample Freezing Platform"*

As regards chemical analysis, “traditional” monitoring data are stored in NORMAN EMPODAT [6] (http://www.norman-network.net/empodat/), while monitoring data from NTS are stored in a newly-designed database, the NORMAN “Digital Sample Freezing Platform” (NORMAN DSFP www.norman-data.eu). The NORMAN DSFP has been developed to enable retrieval of occurrence data of chemical compounds (established NORMAN compounds and new substances) via retrospective analysis of environmental samples analysed with LC-HR-MS in full scan and exported as searchable peak-lists.

The first trials started in 2016 and it is now possible to check for presence/absence of more than 14,000 substances listed in SusDat in any sample by looking at their exact mass (of the most probable adduct ion), in combination with RTI (within its tolerance range). For compounds for which the fragmentation behaviour is known (i.e. present in NORMAN MassBank [7]), fragments can be searched in data-dependent raw data or/and in high collision energy layers of data-independent data. For compounds with unknown fragmentation behavior, *in silico* predicted fragments are generated and can be used for screening purposes. In this way the DSFP allows for screening of presence of virtually any substance detected by powerful LC-HR-MS in any kind of (e.g. environmental) matrix.

NORMAN DSFP is using the harmonised Data Collection Template (DCT), which enables collection of instrumental and sample-specific metadata and allows possible connection with EMPODAT. Moreover, metadata stored in the platform allow for spatial and temporal analysis of presence of these compounds in 'digitally frozen samples'.

For proper storage of NTS data in DSFP the contributors should provide, as basic prerequisites: 1) the “harmonised” DCT filled-in, 2) the chromatograms obtained by the instrumental analysis and 3) the observed experimental retention time of the RTI mixture, injected under the same conditions of the sample. A procedure has been launched to avoid the laborious task of filling in the DCT and make the contribution of NTS data easier.

In the coming years, efforts will be focused on increasing the volume of NTS data digitally-archived in the NORMAN DSFP and on increasing the performance of the software (e.g. with inclusion of GC-APCI-HRMS/GC-EI-HRMS data, semi-quantification of the suspects).

*Exposure Index*

For industrial chemicals an exposure index developed by Stellan Fisher (KEMI) is already available in SusDat. It is based on three different components (independent datasets):

1. The annual tonnage (AT index) calculated using tonnage data available in the REACH database

2. Measure for release during use (use index, UI), derived from the SPIN database

3. Measure for range of use (range of use index, RI), considering, e.g., a wide dispersive use, derived from the SPIN database.

The value of each of the three scores should be between 0 and 1. The final exposure index according to KEMI is calculated as the sum of each sub-score divided by 3.

$$\textit{Exposure Index (KEMI)} = [\, AT + UI + RI \,] \div 3 \quad \textit{(range 0 - 1)}$$

A proposal for the derivation of an exposure index for biocides is enclosed as an annex and was presented by Fh-IME at the WG-1 meeting in June 2016 in Paris. The concept for the derivation of this index is similar to the one developed by KEMI for industrial chemicals.

The maximum score of each component is 5. The total score for the exposure index is calculated as

*Exposure Index (Fh-IME biocides) = (QU score + RDU score + WDU score) ÷ 15*

Where,

1. QU = Quantity Used (based on estimation of the marketed tonnage)

2. RDU = Release During Use (based on a qualitative assessment of the emission relevance)

3. WDU = Wide Dispersive Use (based on the number of registered biocidal products containing a certain biocidal compound).

The final exposure index is a value between 0 and 1.

The exposure indicator can be used as a surrogate for monitoring-based observed exposure in order to allow ranking of compounds in Cat. 2, 4 and 5 for which data are lacking or are of insufficient quality.

For pharmaceuticals a first proposal was presented by INERIS in the WG-1 meeting in June 2016. However, the lack of tonnage data for pharmaceutical active substances is an important barrier.

*Literature MEC database*

In the context of the global SAICM process (http://www.saicm.org/Home/), the German Environment Agency started a comprehensive literature review of papers on measured environmental concentrations of human and veterinary pharmaceutical substances found worldwide and compiled them in a systematic database. This database consists of 631 different pharmaceuticals or transformation products that have been detected in the environment of 71 countries covering all five UN regions.

The concept can easily be applied to other groups of compounds in order to make use of the huge amount of analytical data that is published but not available as raw data (as required in the current prioritisation approach) (see above).

A statistically derived literature $MEC_{max}$ and $MEC_{mean}$ might be derived by taking a weighted average of reported maximum and medium concentrations. These values might be used as risk and ranking indicators to distinguish the priority of compounds in Cat. 2, 4 and 5 for which data are insufficient, completely lacking or of insufficient quality.

An additional application of this information in the categorisation process could be explored. For example, for compounds where no data is available, the reported MEC values (e.g. maximum or mean annual concentrations of a number of sampling sites in one country or river basin) might be the only source of information about the environmental occurrence of an

emerging pollutant, and hence could be used to distinguish between compounds for which we have no information at all, and those with insufficient or even comprehensive data.

For those compounds with yet insufficient data (e.g. Cat 2.), additional evidence from the literature data may allow for an improved classification when the number of countries and number of sites) is increased sufficiently thanks to these additional data source.

# 3 List of substances: from hundreds, to tens of thousands candidate substances, how to organise the categorisation / prioritisation process for a much larger list of compounds

*Current status*
Today, the list of candidate compounds ("universe") for prioritisation is the so-called "NORMAN List of emerging substances", defined and compiled based on expert judgment and literature review. Today the NORMAN List of emerging substances includes about 800 compounds, but NORMAN is now building a much larger list of compounds, compiled from lists of substances present on the NORMAN Suspect Exchange initiative and which form the NORMAN SusDat database.

The aim is that this large list (already more than 40,000 compounds) will become the new "universe" of candidate compounds for prioritisation, and the NORMAN List will be compiled as a list of the top priority compounds in each category.

*Issues and proposals*
- For a great majority of the SusDat list of compounds, most, if not all, of the data required to support any possible decision-making process, are lacking. With the current prioritisation scheme most substances would be allocated to cat 2F, 4F, or 5F (due to lack of monitoring data, appropriate analytical methods or effect data) and it would be difficult to discriminate them using the current prioritisation indicators. It is therefore necessary to integrate additional indicators in the prioritisation scheme, i.e. we need a revision of the categorisation and ranking criteria if we want to be able to deal with tens of thousands of compounds instead of hundreds.
- NTS mass spectral information, especially when combining the digital archives on the NORMAN Digital Sample Freezing Platform (DSFP) with the substances in SusDat, offers some potential for prioritising substances (potentially only isobars) that are frequently detected with NTS methods.
- A new algorithm is needed: 1) to allocate compounds of SusDat to the various action categories and 2) to prioritise the substances within each category based on the focus for further actions (e.g. Cat 2 for monitoring actions, Cat. 3 for improvement of ecotoxicological data, Cat 4 for improvement of analytical methods, etc.).
- Exposure index data (KEMI Exposure Index for industrial chemicals and Fh-IME Exposure Index for biocides) and Measured Environmental Concentration (MEC) data from literature data (Literature data Database) should also be part of the new algorithm for the categorisation / prioritisation process.

*See Section 6 for further discussion about the overall review of the decision tree. Go directly to Section 6.3 and Section 7.1.2 to view proposals for integration of NTS-related data in the substance categorisation and prioritisation process.*

# 4 How to deal with multiple commercial products/CASRN associated with analysed compounds

There are several issues that need to be taken into account when dealing with risk assessment and prioritisation of compounds that correspond to commercial products on the market.

First of all, for many organic chemicals, the commercial form of the substance does not correspond to the organic species to analyse, which is the molecular ion of molecules.

Secondly, a compound relevant for chemical analysis is in general associated to multiple commercial compounds (and thus multiple CASRN). In merging the substances in the NORMAN Suspect Exchange to form SusDat, it is obvious that many sources contribute the same “active ingredient” in various salt forms.

The inventory of CASRN of the commercial compounds (e.g. salts) is necessary for retrieval of information about their tonnage, use pattern, experimentally measured properties (e.g. PNECs), etc. and they affect the exposure index values.

Data may be available for one salt form, but not another – shown below in Figure 1. Sometimes the counterion in the salt may be more toxic than the organic component, for example the nickel salt of citric acid. It is a challenge for prioritisation to take into account all relevant data hidden in the various forms, considering also the analytical methods used. In SusDat, the so-called MS-ready form (middle of Figure 1) is stored with a connection to the original data of one of the forms provided to the Suspect Exchange. A close exchange with US EPA CompTox Chemistry Dashboard (https://comptox.epa.gov/dashboard/) is ongoing to define how this can be solved in a more comprehensive manner (e.g. allowing connection to all possible forms).

Curated parts of the Suspect Exchange are already available as lists on the Dashboard (https://comptox.epa.gov/dashboard/chemical_lists) and connection to the MS-ready forms will be available in a future release.

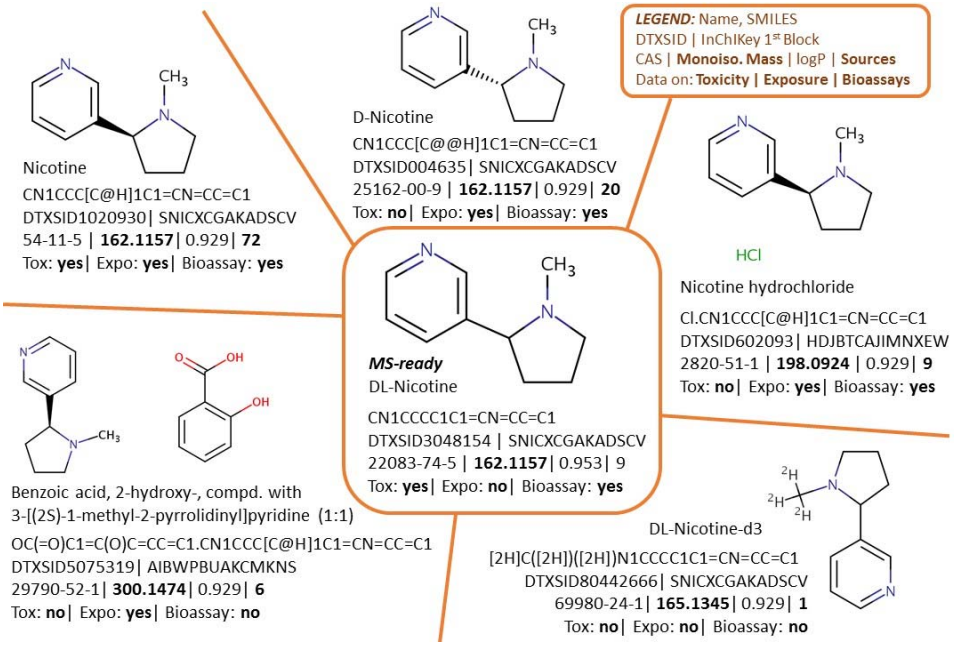



**Figure 1: Various forms of chemicals and the available data (see legend top right) using the example of Nicotine. Source: Schymanski & Williams, 2017, DOI: 10.1021/acs.est.7b01908**

Third and most difficult to address with our common approaches are the Material of Unknown or Variable composition, Complex reaction products or Biological materials (UVCBs) which are a highly challenging case. UVCBs are on the market, many of them with extremely high exposure potential (e.g. through use and tonnage produced).

The components of organic mixtures need to be considered in prioritisation, yet many UVCBs encompass potentially thousands of isomers that cannot be separated with typical analytical methods (and in many cases the UVCB is a technical mixture, so full elucidation of the exact structure(s) is of limited use).

So-called representative structures can be used to define given structures and enable some form of prediction – to couple this with NTS, the example of LAS given below shows that one or two representatives "per mass" are available (see Figure 2 below).

**Questions:**

In terms of monitoring, the "MS-ready" form is the form detected with mass spectrometry. Disentangling how much of the detected form may correspond with which mixture is beyond the scope (and possibilities) of most campaigns. Thus, we may need to deal with "worst case" values over all possible forms, which may result in an overestimation of the hazard for some cases (e.g. Nickel citrate). Do participants agree with this approach?

UVCBs are a highly complex and challenging case, so-called "representative structures" could be used to bring some cases into prioritisation schemes. Some cases such as LAS are being "prototyped" to investigate the effectiveness of various approaches for future discussions. Are participants comfortable with moving forward with this "prototype concept"

(parallel to approaches being trialled in the USA and Canada and in discussion with these partners) to start to assess how to deal with UVCBs in the future?

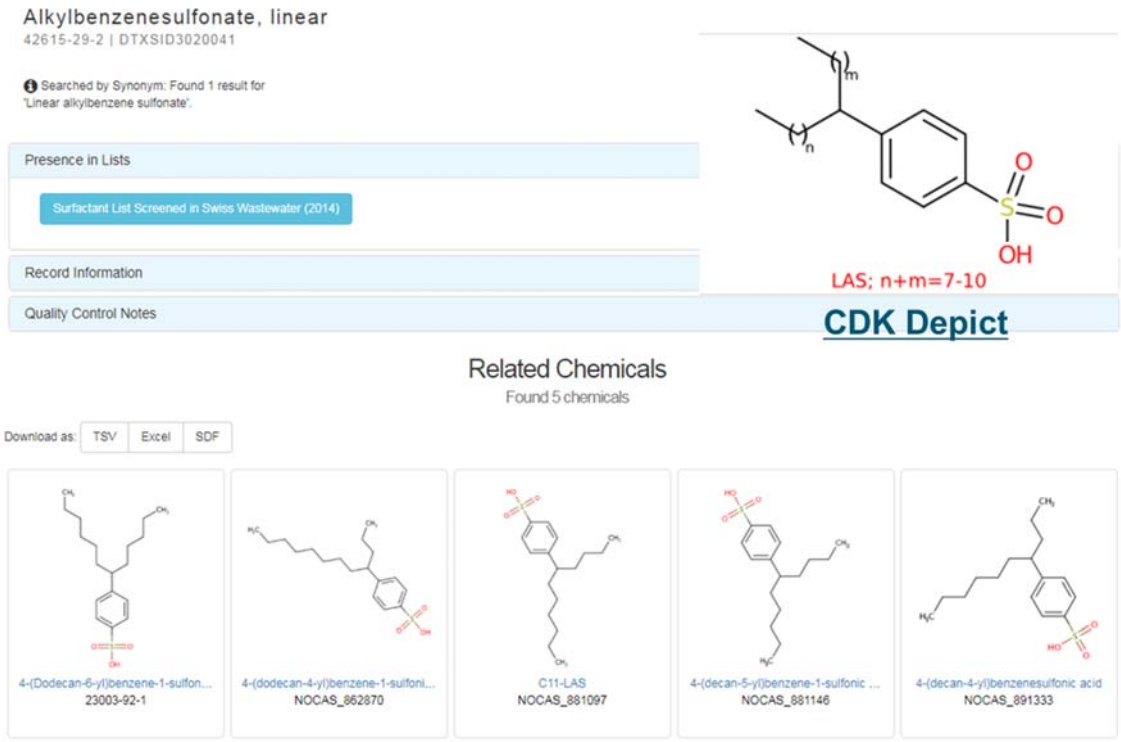



**Figure 2**: **Representative structures to deal with UVCBs: case-study of LAS**

# 5 Monitoring data for prioritisation: how to identify and treat outliers

**NOTE: This section refers to "traditional" target monitoring data collected in existing chemical monitoring databases such as NORMAN EMPODAT and the European platform IPCHEM**.

*Current status*
Today EMPODAT contains about 10 million monitoring data for about 500 substances. No automated procedure is currently in place for identification and treatment of outliers for these monitoring data.

*Proposal*
We propose to adopt the JRC guidelines [ref.] for treatment of outliers with little modification. The definition of outliers is based on the intrinsic distribution of the statistical sample (interquartile ranges).

We suggest a two-levels outlier detection based on different multiples 'k' of the interquartile range. The principle is described in the figure below.

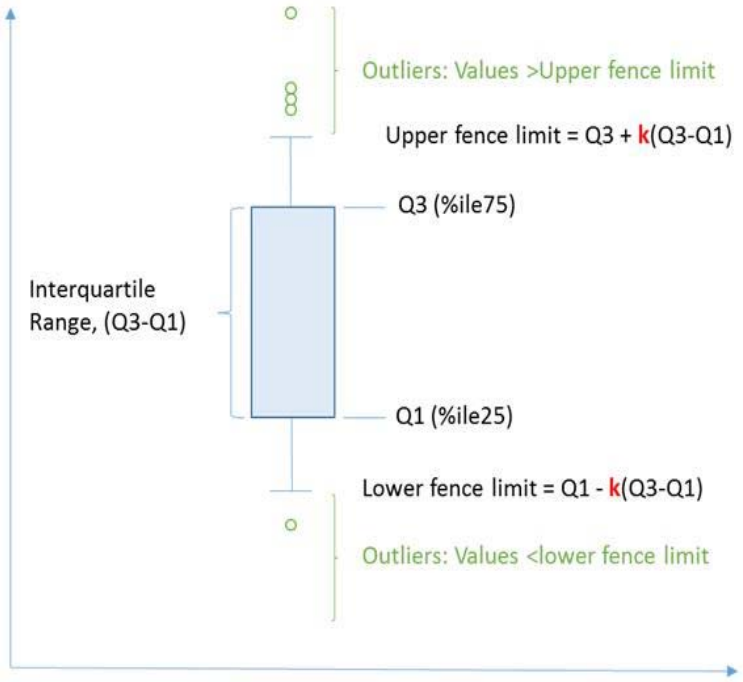



**Figure 3: Scheme for the distribution of measurements in a boxplot and the identification of outliers by defining the fence limits with respect to the interquartile range**

The two action levels could be:
- Higher warning level k=1000
- Low warning level k= 5 – 50

The final action to undertake should be defined by experts, but the default action should be different depending on the type of outlier. If outliers are above the lower warning level, data should be included by default and kept unless it is decided by expert judgment to discard them. If outliers are above the higher warning level, data should be discarded unless they are validated by expert judgment.

**Questions**:
Permanent discarding of datasets identified as outliers should be avoided: discarding of outliers might lead to elimination of concentration data associated with specific important hot spots from industrial emission sources which may be relevant for assessment of priority substances at river basin level. The consequences of elimination of outliers should be discussed. One option could be to implement a procedure (as in the Ecotox module), where experts can “select” the datasets (occurrence data) to be used, analogously to the derivation of PNECs in the Ecotox module. A collaborative procedure could help to assess outliers.

Now that it has been agreed that IPCHEM will replace EMPODAT for collection of monitoring data and NORMAN can retrieve data in IPCHEM (via a dynamic link), can we consider that the treatment of outliers is already done by IPCHEM? What is the data quality check procedure applied by IPCHEM? We propose to get access to all data and do our own outlier treatment according to the procedure described above.

What is the quality of CHEMICAL curation done by IPCHEM? If monitoring data will come under multiple CAS numbers, to what extent will they be inter-connected? How comprehensive will be the work of IPCHEM experts to ensure proper linking of all the connected CAS numbers (associated to all commercial forms of the substance). It is important to discuss this at an early stage since this task represents a lot of work which is very easily underestimated.

We propose to assist JRC in this process and provide them with the curation workflow. The simplest way would be to generate a kind of translation table which would link all CAS-NR to the respective structure identified in NORMAN. This would ensure consistency among the various prioritisation efforts.

# 6 Improving categorisation of the list of candidate substances

This section aims to present options for improvement of the substance categorisation process in the NORMAN framework.

It is worth recalling here that the overall prioritisation procedure is carried out in two successive stages. In the first stage, see Figure 4, a decision tree classifies chemicals into six categories, based on identified ("categories" of) knowledge gaps and actions to be taken by the research community and public authorities to fill them. The second stage entails the ranking of the substances within each (action) category, on the basis of the criteria / indicators identified for each category.

The overall process is an iterative one that involves a periodic revision of the priority substances in each category whenever e.g. new information / more reliable data are generated or feedback from applied reduction measures is available.

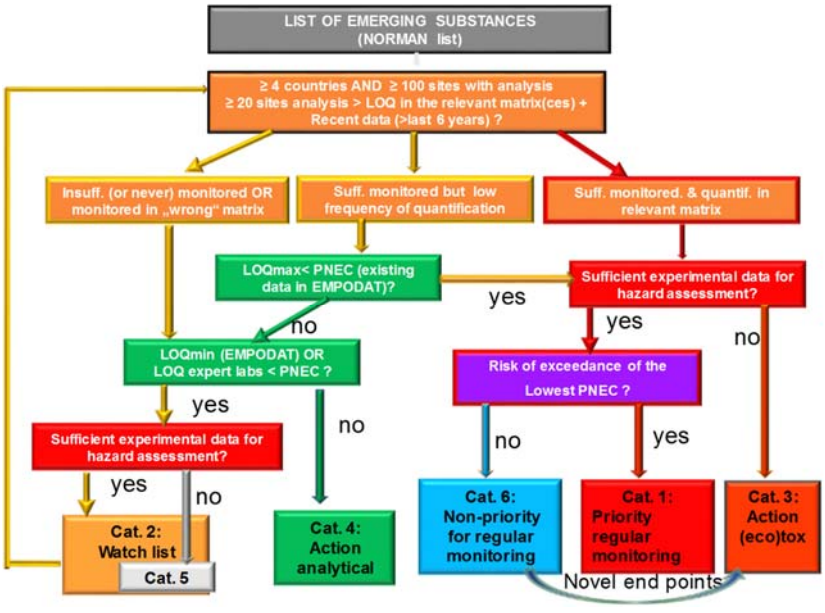



**Figure 4: NORMAN decision tree for categorisation of substances**

The first step in the decision tree consists of grouping the compounds by degree of investigation and evidence of exposure.

On the basis of a set of indicators (see **Table 1**), the candidate substances are divided into distinct groups:
- Substances that are **sufficiently monitored** and **sufficiently quantified** in the relevant matrix;
- Substances that are **sufficiently monitored** in the relevant matrix, **but with a low frequency of quantification**;
- Substances for which we have **no or insufficient data** in the EMPODAT database or other existing datasets (labelled as "never monitored").

In the light of the considerations made in the previous sections about the need to integrate new criteria for categorisation of substances, we can assume that **for substances for which we already have sufficient monitoring data we do not need to study the possible integration of new indicators associated with NTS data, literature data etc.**

Based on this assumption, in the following sections (Section 6.1 and Section 6.2) we discuss proposals for improvement of categorisation criteria for the "sufficiently monitored" compounds, with reference to target monitoring data (already available in EMPODAT), only.

Options for integration of indicators associated with the **MS data in the categorisation process are proposed in Section 6.3 for substances for which no or insufficient monitoring data are available in the EMPODAT database or other existing datasets**.

## 6.1 Reviewing the criteria to define "sufficiently" monitored / "sufficiently" quantified substances

Here we address the definition of the terms:
- "sufficiently / insufficiently monitored"
- "sufficiently / insufficiently quantified".

We need to revise the criteria to define these two terms. Some proposals are provided in the sections below.

NOTE: The criteria are different depending on the geographical scale of the prioritisation study.

### *6.1.1 Application at the EU level*

*Current status*
The criteria and cut-off values associated with the different indicators for exposure assessment are summarised in the table below.

**Table 1: Cut-off values associated with the different indicators used for exposure assessment in the categorisation process**

| Indicators / Substances sub-groups | Analyses available in the relevant matrix(ces) | Number of countries with analyses | Number of sites with analyses | Number of sites with analyses > LOQ |
|---|---|---|---|---|
| **Subst. suff. monitored and sufficiently quantif. in relevant matrix** | Yes | ≥4 countries | ≥100 sites | ≥20 sites |
| **Subst. suff. monitored but with low frequency of quantification** | Yes | ≥4 countries | ≥100sites | <20 sites |
| **Subst. insufficiently monitored** | Yes | <4 countries AND / OR <100 sites with analyses | | Not relevant |
| **Subst. never monitored (i.e. data not available in EMPODAT or other existing datasets)** | Not relevant | No data | No data | No data |
| **Subst. monitored in a "not relevant" matrix** | No | Not relevant | Not relevant | Not relevant |

*Proposals*

- Number of countries and Number of sites with analyses: should we revise the current cut-off values? Based on the results of the 1st EU Watch list monitoring report, "100 sites" seems to be considered by JRC as the reference cut-off value for sufficiently monitored substances at EU level. However, with the future development of the EU Watch List and the increase in size of the IPCHEM database these cut-off values could be revised to be more in line with the average number of available monitoring data.
- 50 sites with analyses > LOQ instead of 20; 50 sites enable a more robust MEC95 indicator against outliers. Moreover, *50 sites ensure that the evaluation of the risk ratio is performed on a sufficiently significant number of sites (at least 4 sites with exceedance if we have 50 quantified sites)*
- *Delete the "relevant matrix" criterion?* If data are available they are anyway useful, even if they are not in the relevant matrix. On the other hand monitoring data in the "wrong" matrix may lead to underestimation of the risk (for example, for very hydrophobic compounds, measurement in water may lead to low frequency of quantification, whereas there might in fact be a risk when analysing biota).

### *6.1.2 Application at the national level*

*Proposals*

The "sufficiently monitored" criterion should be defined at the level of a country based on the requirements for "sufficient" level of monitoring in the national regulation. In France, a compound is considered as "sufficiently monitored" when it is monitored at 20% of the stations of the regular monitoring network (about 1500 stations). This is the monitoring level applied for the substances on the French Watch List.

Based on these principles, we propose as indicative cut-off values for "sufficiently monitored" substances at the national level:

- 1/3 river basins (2 in France)
- 20% of the stations of the regular monitoring network (300 in France)

For the "sufficiently quantified" criterion at the national level:

- 20 – 50 stations with a risk evaluation calculated with MEC95 (necessary for sufficiently robust MEC95) for countries with a large number of stations;
- 10 stations with a risk evaluation calculated with MEC90 for countries with lower monitoring coverage.

## 6.2 Refining the criteria for categorisation of substances with sufficient monitoring data

### *6.2.1 Revising criteria for categorisation of substances with sufficient monitoring data but low frequency of quantification*

*Current status*
In the categorisation algorithm, the key criterion to categorise well-monitored but not sufficiently quantified substances in Categories 1,3 or 6 OR 2,4 or 5 is: "$LOQ_{max}$ >PNEC?". This means that we are requiring all data to have limits of quantification compatible with PNEC.

NOTE: Current use of LOQ90 instead of LOQmax should be checked.

*Critical Issues*
$LOQ_{max}$ > PNEC appears to be excessively selective as a criterion: a compound measured a thousand times, but quantified less than 50 times, with one very "high" LOQ (leading to $LOQ_{max}$>PNEC), would be placed in Category 2 thereby leading to further monitoring while the substance could probably be better placed in Category 1 (if there is a local risk), OR in Category 6.

*Proposal*
Substances that are sufficiently monitored (at least 4 countries and at least 100 stations), with low level of quantification (< 50 sites with data above PNEC) should not be classified in Category 2, if the dataset contains a "sufficient number of stations" with "good quality" LOQs (i.e. stations with $LOQ_{min}$ lower than PNEC). We can define as "sufficient number of stations" the number of stations required for "sufficiently monitored" substances, for example 100 stations for the EU Watch List. In conclusion, this refined criterion allows substances characterised by a low level of quantification but already well monitored and with enough "good quality" data to be classified in Categories 1,3 or 6, instead of Category 2.

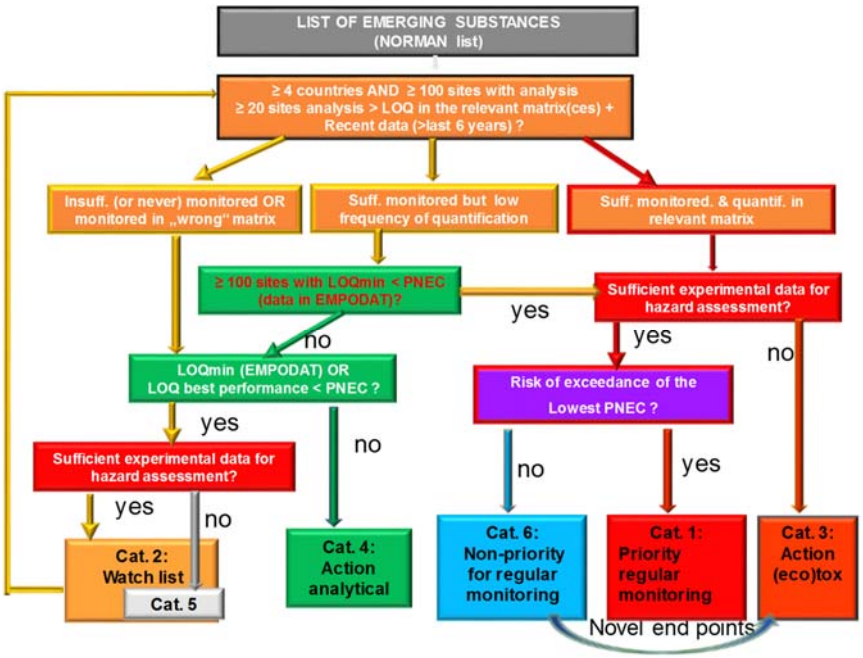



**Figure 5: New criterion proposed for categorisation of well-monitored substances with low frequency of quantification between categories 1,3, 6 AND 2,4, 5**

### *6.2.2 Revising the distribution of well-monitored substances between Categories 1 and 6*

*Current status*

For substances with sufficient monitoring data and quantifications, the criterion used for assessment of risk potential is “MEC95/PNEC >1?”. If MEC95/PNEC is >1, the substance is classified in Category 1 (1A), if not it is assigned to Category 6 (6A).

For substances sufficiently monitored but lacking quantifications it is still possible to allocate these substances to Category 1 if the potential for a local risk is identified (Cat 1B: priority substances for control measures at local level). The criterion used in the current scheme is “$MEC_{site\_max}$/PNEC >1?”. Depending on this test, the compound is classified in categories 1 (1B) or 6.

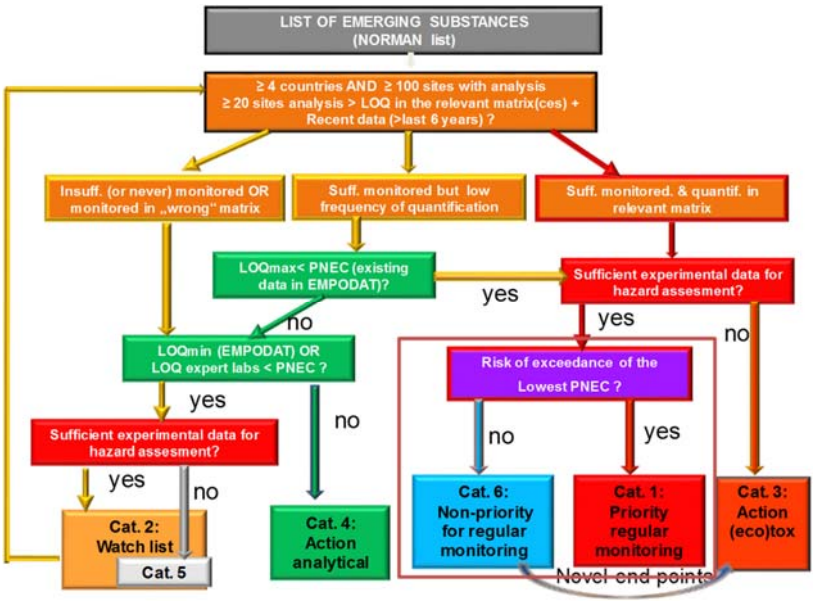



**Figure 6: Current scheme for categorisation of substances between categories 1 and 6**

*Critical issues*

The use of this differentiated criterion has some limitations and may lead to inconsistencies for further actions.

For well quantified compounds, it is necessary to have 5% of stations with quantified data above PNEC to classify a substance in category 1A. However, this could lead to allocation of possible problematic substances to Category 6A instead of Category 1A. For instance, a well-monitored compound (2000 stations) and highly quantified (1000 stations) would be assigned to "Category 6" if less than 50 stations have data above PNEC. However, a substance quantified at 40 stations with concentrations above PNEC, especially with high exceedances, should not be considered as a low priority substance.

On the contrary, substances lacking quantification may be considered priority compounds (in category 1B), whereas they should be allocated to category 6. For instance, for a compound with a 1000 sites with monitoring data, but only 10 quantifications, 1 station with $MEC_{site}$ above PNEC would be sufficient in the current scheme to classify the substance in Category 1B (local risk).

*Proposals*

Better harmonisation of these criteria is necessary. We propose a double criterion for assessment of risk potential: first the MEC95/PNEC criterion should be checked for a well quantified compound, and, if this one is not fulfilled we propose to compare $MEC_{25th}$ ($25^{th}$ highest $MEC_{site}$, where $MEC_{site}$ is the maximum concentration at each site) to the PNEC. The new Category 1B corresponds to compounds fulfilling the $2^{nd}$ criterion. If they fulfil neither of them, these compounds are classified in Category 6.

Regarding compounds lacking quantification (less than 50 sites > LOQ), $MEC_{95}$ cannot be derived (due to insufficient data). Therefore, these compounds should only be tested against the second criterion and classified as 1B or 6.

The algorithm proposed is described in the figure below.

The cut-off value (25 stations) is considered representative of a non-negligible risk for the environment. However, this number should be open for debate. The key question for its definition and acceptability is the following: What is the number of stations with exceedance of PNEC that can be considered as significant for decision-makers to trigger actions?

The proposed double criterion would lead to a new categorisation of substances, as shown in the figure below.

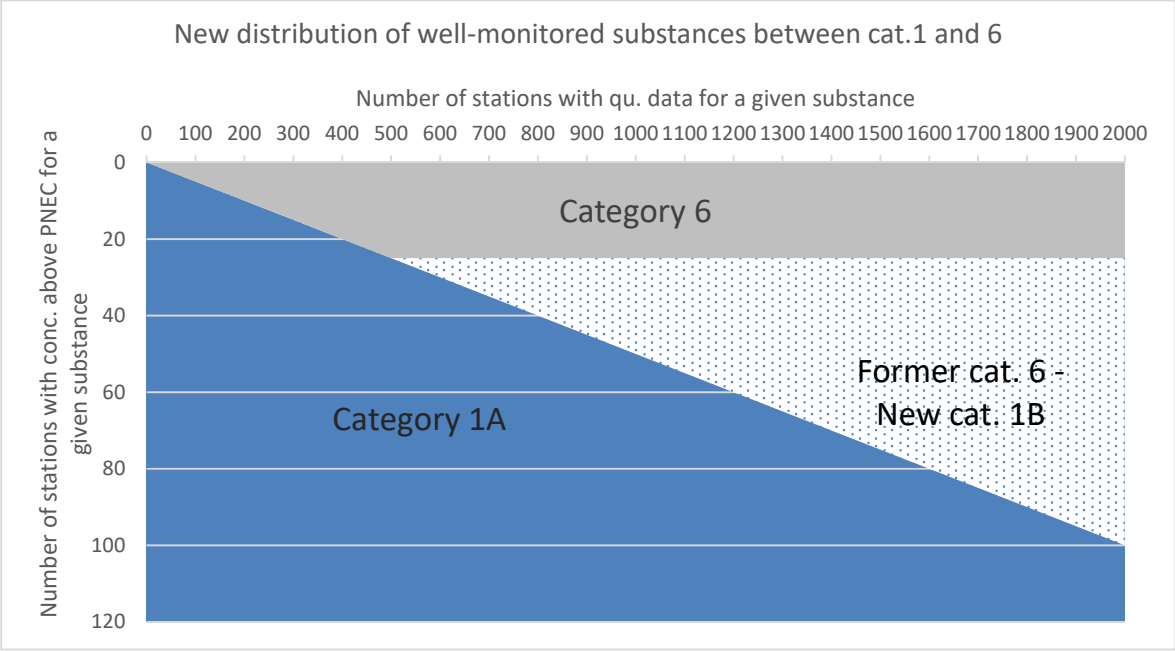



**Figure 7: Revision of distribution of substances between Categories 1 and 6**

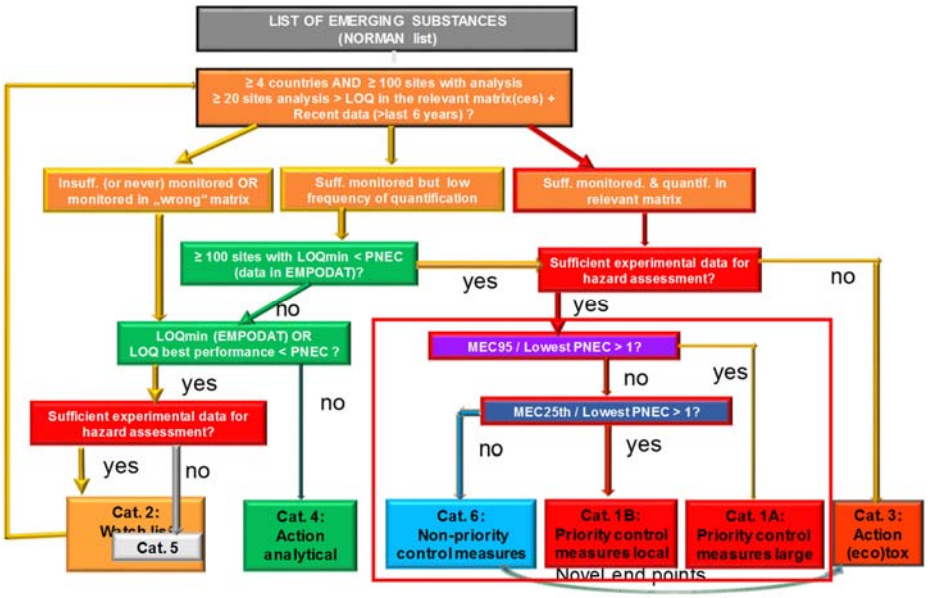



**Figure 8: New proposed scheme for categorisation of substances with reviewed criteria for distribution of substances between Categories 1A, 1B and 6**

### *6.2.3 For better accounting of the relevance of Cat 1 substances: European or national relevance*

*Current status and critical issues*

The experience from past prioritisation runs (in particular the one carried out to give recommendations to the JRC for the review of the list of PS) we have started to propose harmonised requirements as regards the definition of a Priority Substance at EU level.

In the NORMAN recommendation document of September 2016 for the review of the WFD PS we wrote that a substance can be considered as relevant at EU level when PNEC exceedance is observed in more than 3 countries where the substance has been recently monitored (after 2008) or in at least 50% of the investigated countries (always considering recent data, after 2008).

However, this definition is not yet formally accepted by NORMAN WG-1 and not yet included in the prioritisation algorithm.

We also noticed in the last prioritisation exercise for the Commission that it is often the case that a substance which would fulfil globally the requirements for Cat. 1, would show PNEC exceedance only in a limited number of countries.

However, when looking in more detail in the datasets it was obvious that this was due to insufficient quality of the available monitoring data in some of the countries. In the countries where the substance was measured with a satisfactory LOQ, PNEC exceedance was frequently observed.

This was the case for fipronil and triclosan, which were anyway proposed as candidate PS based on the consideration that when discarding the countries with insufficient analytical performance (LOQ > PNEC) PNEC exceedance was confirmed for 50% of the countries, although the number of countries would then be relatively low (i.e. 2 out of 4 countries with good quality data) when considering that recent data were available in the DB for 9 countries.

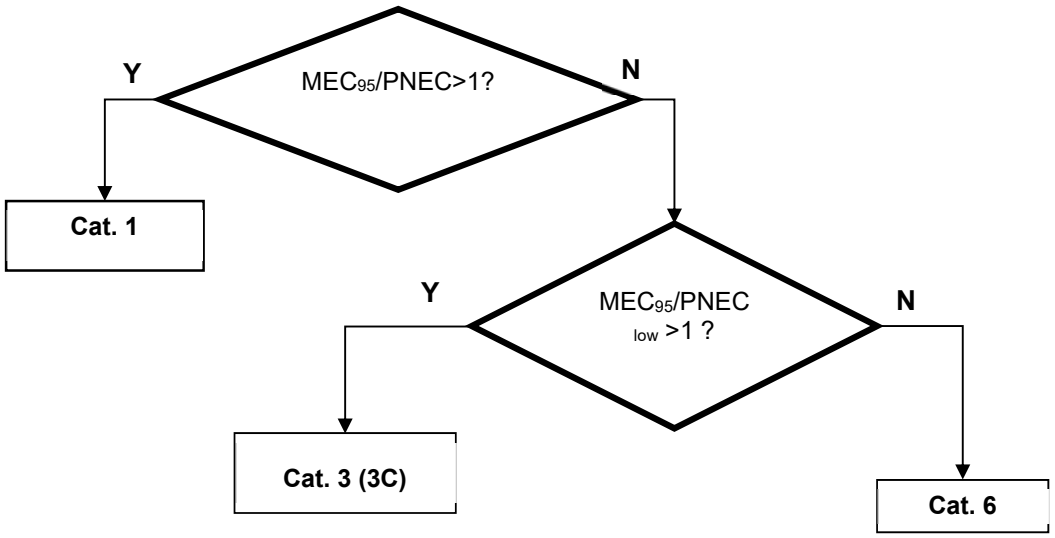
*Proposal*
We want to propose to show in Cat 1A the number of countries for which exceedance of PNEC is observed and the number of countries where poor quality data could have biased the final results / risk ratio estimation. In this we will have two sub-categories: 1A_EU and 1A_national. We would like to put up for consideration the proposal that PNEC exceedance in 3 countries with recent data OR 50% of the investigated countries with good quality data ($LOQ_{min\_country}$ < PNEC) should be sufficient justification for considering the European relevance of a compound.

### *6.2.4 For better accounting of PNEC uncertainties: refining criteria for Category 3*

*Current status*
In spite of the safety factors applied in the derivation of the PNEC values, we can argue that uncertainties are not fully considered.

*Proposal*
We propose to compare not only $MEC_{95}$ to the Lowest PNEC chosen by experts, but also to compare $MEC_{95}$ to the lower extremity of the PNEC uncertainty interval, or to compare it with the lowest available value of PNEC available in the Ecotox module. This might allow us to flag up some substances in Category 6 which might require further experts' discussions on the (eco)toxicity of the molecule. A sub-category in Category 3 (3C) could be created to classify these molecules.

**Figure 9: New algorithm proposed to determine between Cat 1, 3, and 6**



This would allow the re-categorisation of substances as shown in the figure below.

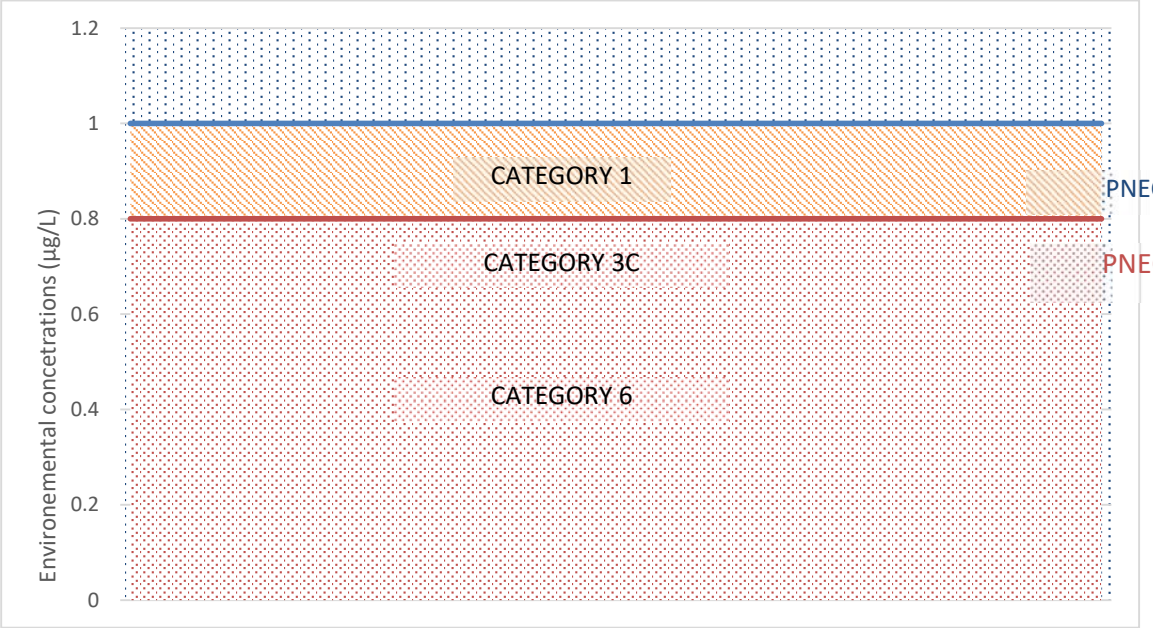



**Figure 10: Example of a distribution leading to new categorisation of the substance**

## 6.3 New criteria for categorisation of insufficiently monitored compounds, taking into account NTS data and other information sources

*Current status*
There are not yet defined criteria for allocation of the list of candidate substances (list of compounds compiled in SusDat) to action categories based on information other than "traditional" target monitoring data available in EMPODAT.

*Proposal*
As a first improvement, we propose to use information derived from retrospective analysis of HRMS data through NORMAN DSFP as additional supporting information for identification of priority substances. This means that we need to **refresh the algorithm of the categorisation scheme** in order to allow the categorisation of SusDat substances based on the qualitative and semi-quantitative information that can be derived from NTS data.

A proposal for revision of the NORMAN decision tree is illustrated in Figure 11 (cf. also Table 2 to Table 5 below).

In this new version of the decision tree we propose to insert a new query ("**NTS data with sufficient quality?**" in Figure 11) which means that, for substances that are insufficiently monitored (or sufficiently monitored but with insufficient "good" quality data) we check whether we have or do not have sufficient identification evidence to claim that the identification of the compound from digitally-archived data is unambiguous.

**The compounds that are unambiguously identified can proceed to the next steps towards Cat 2 / Cat 5 sub-categories, whereas the compounds for which identification is not proven with sufficient confidence will be allocated to Cat. 4.**

As regards **Cat 2 and Cat 5**, after the identification step, we propose to introduce in the decision tree some new queries in order to allow allocation of the substances to sub-categories 2 and 5 ($A^{++}$, $A^{+}$, $A^{-}$ and $B^{++}$, $B^{+}$ and $B^{-}$) based on an estimate of their spatial occurrence and potential risk of exceedance of environmental quality criteria (PNEC or P-PNEC).

The indicators could be expressed in terms of

- Nb. countries (*x% of countries)* with positive detections
- Nb. sites (or *x% of sites)* with positive detections
- Nb. sites with exceedance of PNEC (or % of sites with exceedance of PNEC)
- Extent of exceedance of PNEC.

For the **substances in the new Cat. 4**, the action will address not only the improvement of the LOQ (as in the "old" scheme) but also the compound identification aspects. In this context, we should also address here "real" NTS data, i.e. peaks where we know only m/z, RT, and optionally MS/MS spectra, but the structure is not yet elucidated.

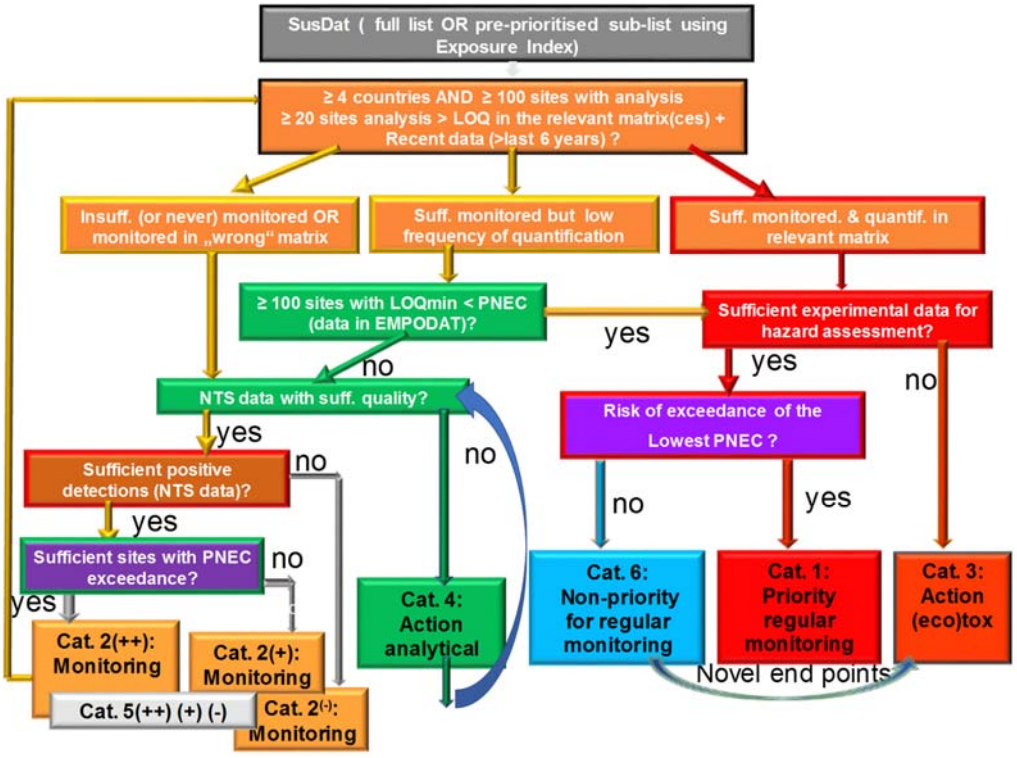



**Figure 11:** Proposal for revision of the NORMAN decision tree for categorisation of substances

## Category 1

| Category | | Current scheme with target monitoring data | Proposal for revised scheme | |
|---|---|---|---|---|
| 1 | 1A | Sufficiently monitored and sufficiently quantified substances for which a risk is identified | 1A | Proposals for update of the criteria for categorisation have been presented earlier (see section 6.2.2).<br>In addition to these proposals, new criteria could be introduced to integrate NTS data from retrospective analysis for identification of Cat 1 compounds. For example, we could have substances for which a lower number of sites with quantified monitoring data (from target analysis) is available but which could be still be allocated to Category 1 as a result of evidence from retrospective analysis NTS data. The criteria for this new Category 1 are to be discussed. |
| | 1B | Sufficiently monitored substances, with a low level of quantification, but for which a risk is identified at the local level (i.e. $MEC_{site\_max}$ > Lowest PNEC) | 1B | Proposals for update of the criteria for categorisation of substances in Cat 1B have been presented earlier (see section 6.2.2) |

**Table 2:** Proposal for revision of the categories in the NORMAN scheme – Category 1

## Category 2

In the current version of the prioritisation framework substances are allocated to Category 2 when monitoring data are either insufficient or of insufficient quality.

There are two sub-categories for Category 2. Category 2A contains substances with insufficient or no monitoring data. Category 2B contains substances with sufficient monitoring data but for which the frequency of quantification appears to be low and there is evidence of poor quality data which could explain the low frequency of quantification of these substances.

In the new scheme, NTS data could be used as an additional source of information to better discriminate the actions needed for these substances and their level of priority within each category.

We propose in the table below to convert Cat 2A of the "old" scheme to the new sub-categories $2A^{++}$ $2A^{+}$ and $2A^{-}$ (and the same for Cat 2B) based on:

- Frequency of detection
- Frequency of exceedance of the PNEC.

For allocation of substances to Category 2, the criteria for "unambiguous identification" of the compound must in any case be fulfilled.

| Category | | Current scheme with target monitoring data | Proposal for revised scheme | |
|---|---|---|---|---|
| 2 | 2A | Insufficiently monitored substances for which further monitoring data are needed | 2A | Substances with insufficient target monitoring data AND NTS data with "sufficient" identification proof |
| | | | | 2A(++)<br>- *sufficient positive detections*[1] from NTS retrospective analysis |

[1] > x countries and / or > y sites, with x and y to be defined

| | | | | |
|---|---|---|---|---|
| | | | | - AND with > x% (x = sufficient) of sites exceeding the Lowest PNEC value |
| | | | | Cat. 2A(+)<br>- sufficient positive detections from NTS retrospective analysis |
| | | | | Cat. 2A(-)<br>- *insufficient positive detections*[2] NTS retrospective analysis |
| | 2B | Sufficiently monitored substances, with a low level of quantification and poor quality data (target monitoring data), further monitoring data are needed | 2B | Substances with sufficient monitoring data, with a *low level of quantification*[3] and *poor quality data*[4] (target monitoring data), AND NTS data with "sufficient" identification proof |
| | | | | Cat. 2B(++)<br>- *sufficient* positive detections from NTS retrospective analysis<br>- AND with > x% (x = sufficient) of sites exceeding the Lowest PNEC value |
| | | | | Cat. 2B(+)<br>- *sufficient* positive detections from NTS retrospective analysis |
| | | | | Cat. 2B(-)<br>- *insufficient* positive detections NTS retrospective analysis |
| | 2F | No occurrence data are available in EMPODAT (or other datasets) but the literature data show that the LOQs associated with existing analytical methods are lower than the Lowest PNEC | 2F | This category can be deleted because if we have *insufficient or no monitoring the substance* will belong to Cat. 2A++ / 2A+ and 2A(-) already mentioned above depending on the results of the NTS retrospective analysis |

**Table 3**: Proposal for revision of the categories in the NORMAN scheme – Category 2

### Category 4

In the "old" scheme, there are three different sub-categories 4 depending whether we have sufficient or insufficient monitoring data (target monitoring data) for these substances. In both cases, monitoring data show that the analytical performance need to be improved ($LOQ_{min}$ associated with current analytical methods are above the Lowest PNEC).

In the revised version of the categorisation scheme, it is proposed to add to the original Category 4 the compounds for which NTS data do not fulfil the requirements for a sufficient level of confidence for compound identification.

| Category | | Current scheme with target monitoring data | Proposal for revised scheme | |
|---|---|---|---|---|
| 4 | 4A | Insufficiently monitored substances for which analytical methods need to be improved (LOQs associated with current analytical methods are above the Lowest PNEC) | 4A | Insufficiently monitored compounds for which NTS data do not fulfil the requirements for a sufficient level of confidence for compound identification.<br>Within this category it will be necessary to use criteria / indicators to identify the substances occurring more OFTEN and with HIGHER INTENSITY. |

[2] < x countries and / or < y sites, with x and y to be defined
[3] > 4 countries and/ or > 100 sites with target monitoring data AND < 50 sites with data > LOQ
[4] < 100 sites with LOQmin < PNEC

| | | | | |
|---|---|---|---|---|
| | | | | We need also to discuss whether it is necessary to group the substances according to these indicators and thus create new sub-categories within Category 4. |
| | 4B | Sufficiently monitored substances, with low level of quantification, for which analytical methods need to be improved (LOQs associated with current analytical methods are above the Lowest PNEC) | 4B | To be completed based on the same considerations as above |
| | 4F | No monitoring data are available in EMPODAT (or other datasets) and no LOQ data retrieved from the literature to define whether existing analytical methods are compatible or not with the Lowest PNEC,<br>OR<br>Monitoring data available in EMPODAT show that the LOQs associated with the available data are above the Lowest PNEC BUT no LOQ data have been retrieved from the literature to define whether the LOQs associated with current analytical methods are above or below the Lowest PNEC | 4F | The "old" Category 4F could be converted to a new Category 4F where we will address real NTS data, i.e. mass hits for which we know only m/z, RTI, etc. but not yet the structure of the compound.<br>NOTE, however, that at the moment in the decision tree there is no clear entry for real NTS data, i.e. the proposed decision tree covers only retrospective screening of suspect substances in full scan HR mass spectra. |

**Table 4:** Proposal for revision of the categories in the NORMAN scheme – Category 4

### Category 5

For category 5 we propose the same scheme as for Category 2. The only difference with respect to Cat. 2 is that the Lowest PNEC values will be here P-PNEC values (modelled data).

In conclusion Category 5 in the current prioritisation scheme can be converted to Cat. 5A(+), cat 5A(++) and cat 5A(-) based on the following criteria:
- Frequency of detection AND
- Frequency of exceedance of the PNEC.

For allocation of substances to Category 5, the criteria for "unambiguous identification" of the compound must be fulfilled.

Cat. 5F: This category is redundant with the other categories mentioned above. It can be deleted.

### Category 6

| Category | | Current scheme with target monitoring data | Proposal for revised scheme | |
|---|---|---|---|---|
| 6 | 6A | Sufficiently monitored and sufficiently quantified substances, with experimental ecotoxicity data, but no risk is identified | 6 | Revision of the categorisation criteria using target monitoring data (see section 6.2.2.) |
| | 6B | Sufficiently monitored substances, with low level of quantification, AND LOQs < Lowest PNEC AND no risk is identified (either at wide or at local level i.e. $MEC_{site\_max}$ < Lowest PNEC) | | |

**Table 5:** Proposal for revision of the categories in the NORMAN scheme – Category 6

**Questions**

Do you agree with the main principles behind the proposal for a “new” decision tree for substances categorisation and the “new” categories (and sub-categories)?

The indicators and the cut-off values for the allocation of the substances to the various categories need to be defined and agreed within the WG. What are your proposals?

Screening of thousands of candidates in NTS data may result in false positive detection hits. Sufficient evidence of identification of a compound from retrospective analysis of full scan digitally-archived HRMS data is a key principle introduced in the decision tree. Possible criteria for “sufficient” identification of compounds (Category 2 vs Category 4), i.e. mass accuracy, plausible RT, isotopic profile, detected fragments, experimental or predicted fragments, need to be discussed.

Quantification is very uncertain for retrospective data, almost impossible when no standard is available. However, semi-quantification is feasible (provided that more parameters are requested per digitalised sample) and Lowest PNEC (or P-PNEC) are available. The semi-quantified PEC / PNEC ratio may be uncertain but contains some information and could be used for implementation of indicators such as “frequency of exceedance of the PNEC” (FoE) and Extent of exceedance of the PNEC (EoE)? Do you consider the derivation of semi-quantified data a possible acceptable solution?

# 7 Improving the ranking algorithm

Once the substances have been allocated to the various action categories, a subsequent ranking of the substances within each action category takes place.

This section describes the proposals for revision of the procedure for ranking the substances within each action category.

NOTE: Since the objectives differ from one category to another (e.g. Category 4 for improvement of analytical performance; Category 3 for improvement of toxicity data), the prioritisation indicators may differ from one category to another as well.

## 7.1 Exposure score

### *7.1.1 Revising the "Exposure score" (monitoring data) for Categories 1,3,6*

*Current status*
As described in the NORMAN prioritisation framework, for Categories 1,3 and 6 the exposure scores are computed as an arithmetic mean of an exposure score based on available monitoring data ("Observed Exposure") and the "Predicted Exposure" score based on an exposure index.

The current "Observed Exposure" score is calculated as the average of the following indicators (see equation below):
- Frequency of observations with concentration >LOQ
- N° of countries with concentration >LOQ
- N° of sites with concentration >LOQ
- Concentration trend (positive / negative trend)
- Observations in groundwater (Yes / No)

*Observed Exposure* = [(score "Freq. observations > LOQ") + (score "No. countries > LOQ") + (score "No. sites > LOQ") + (score "Conc. Trend") + (score "Observation in GW")] / 5

The following table provides further details about the scales for the various indicators.

| Score | Indicators | Description | Formula |
|---|---|---|---|
| Observed Exposure (monitoring data) | A) Frequency of observations with concentration >LOQ | Fraction of analyses >LOQ | = value as a decimal number rounded to two decimals |
| | B) N° of countries with concentration >LOQ | No. of countries with concentr. >LOQ | Value between 0 and 1<br>0 countries (or no data) = 0<br>≥1 country = 0.10<br>≥ 2 countries = 0.20<br>≥ 5 countries = 0.50<br>≥ 10 countries = 1 |
| | C) N° of sites with concentration >LoQ | No. of sites with concentration >LOQ | Value between 0 and 1<br>0 sites (or no data) = 0<br>≥1 site = 0.10<br>≥ 10 sites = 0.20<br>≥ 100 sites = 0.50<br>≥ 1000 sites = 1 |

| Score | Indicators | Description | Formula |
|---|---|---|---|
| | | | |
| | D) Concentration trend | Trend Regression of $MEC_{95/a}$ for > 5 years and > 6 sites | Significant positive trend = 1<br>Positive trend = 0.5<br>No trend = 0.25<br>No data = 0.1<br>Negative trend = 0 |
| | E) Observation in groundwater | Yes = 1<br>No = 0 | = value |

*Critical issues identified & proposals:*

- *Issues when using "Predicted Exposure" scores for Categories 1,3, and 6*

  The predicted exposure scores have limited commensurability across use categories and lack of predicting power. It seems that using them for Categories 1,3 and 6 leads to a decreased robustness of the exposure score. We suggest using the Observed Exposure score, only (derived from monitoring data) for Categories 1,3, and 6.

- *Lack of consistency between "frequency-based" and "absolute value-based" indicators.*

  Both frequency-based indicators and absolute value-based indicators are currently used for the calculation of the exposure score. However, the absolute values are biased by the level of monitoring of the substance, which results in systematically higher scores for the compounds that are more extensively monitored.

  Example: A compound investigated and quantified in 4 countries and 100 stations at 100% of quantification would have a score of : A+B+C = 1+0.2+0.5 = 1.7.
  Another substance (e.g. BPA) quantified on 35% of the analyses in 11 countries and at more than 1000 sites would have a score of A+B+C =0.35+1+1=2.35.
  The first compound is more critical in terms of environmental occurrence, but its score is affected by the lower monitoring of the compound (even if both compounds fulfill the criteria for sufficiently-monitored compounds, i.e. criteria for Cat. 1,3,6).

  We propose to replace "B/Number of countries with quantification" by "Number of countries with quantification divided by total number of countries with monitoring data". Similarly, we suggest replacing "C/ N° of sites with concentration >LoQ" by "Number of sites with quantification divided by total number of sites with monitoring data".

- *Weight of each sub-indicator in the formula for calculation of the "Observed Exposure" score*

  Each indicator has the same weight in the current algorithm. However, the "country indicator" appears as too sensitive to the level of monitoring in the different European countries. The intensity and frequency of monitoring varies significantly between the different countries in Europe. For instance, according to the results available in EMPODAT it appears that BPA is monitored in 11 countries but 95% of the total monitored stations are in 3 countries. In the remaining 8 countries, the limited number of stations with analyses

makes the score very dependent on the quality of these analyses and the representativeness of the sites chosen. To avoid the bias associated with an inhomogeneous level of monitoring on the final exposure score, a lower weight should be given to the "country indicator" (Number of countries with quantification divided by total number of countries with monitoring data).

We therefore propose an exposure score calculated as described below:

20% "Country indicator"(A) + 40% "Sites indicator"(B) + 40% "Analysis indicator"(C)

- *"D/Concentration trend" and "E/Observation in groundwater"*

The "*D/Concentration trend*" and "*E/Observation in groundwater*" indicators of the "Observed Exposure" score are too dependent on the availability of data. The consequence is that they lower the overall "Observed Exposure" score for substances for which no data is available (i.e. no monitoring data in groundwater, not enough data for trend derivation).

Two options are suggested:

1. Delete "*D/Concentration trend*" and "*E/Observation in groundwater*" in the calculation of the final "Observed Exposure" score
2. Use "*D/Concentration trend*" and "*E/Observation in groundwater*" as corrective factors of the "Observed Exposure" score based on A+B+C.
   Regarding the *Concentration trend* score, we propose to keep the current indicator as it is but improve it with consideration about the regulatory status of the substances. We propose to change the scoring algorithm as follows:
   - Significant positive trend, then $CF_{trend}$= 1,2
   - Positive trend, then $CF_{trend}$ = 1,1
   - No trend or no data, then $CF_{trend}$ = 1
   - Negative trend or substance banned or to be banned, then $CF_{trend}$ = 0,9

   Regarding the *Observation in groundwater* score, we propose to keep the current indicator and change the scoring algorithm as follows:
   - Yes, then $CF_{gw}$ = 1,1
   - No, then $CF_{gw}$ = 1

The correcting factors should be included in the calculation of the final „Observed Exposure" score as follow :

***EXPO = (A+B+C)* $CF_{trend}$* $CF_{gw}$***

*Final proposal for "Observed Exposure" score derivation:*

| Score | Criterion | Description | Formula | Final score |
|---|---|---|---|---|
| Observed Exposure (monitoring data) | A) N° of countries with concentration >LOQ / Total N° of countries with analysis | N° of countries with concentration >LOQ / Total N° of countries with analysis | = value as a decimal number rounded to two decimals | **EXPO = (0.2A+0.4B+0.4C) *$CF_{trend}$*$CF_{gw}$** |
| | B) N° of sites with concentration >LOQ / Total N° of sites with analysis | N° of sites with concentration >LOQ / Total N° of sites with analysis | = value as a decimal number rounded to two decimals | |

| | | | | |
|---|---|---|---|---|
| | C) Frequency of observations with concentration >LOQ | Fraction of analyses >LOQ | = value as a decimal number rounded to two decimals | |
| | $CF_{trend}$) Corrective factor for concentration trends | Trend Regression of $MEC_{95/a}$ for > 5 years and > 6 sites + changes in regulatory status | Significant positive trend Then $CF_{trend}$= 1.2<br>Positive trend, Then $CF_{trend}$ = 1.1<br>No trend or no data, Then $CF_{trend}$ = 1<br>Negative trend or substance banned or to be banned, Then $CF_{trend}$ = 0,9 | |
| | $CF_{gw}$) Corrective factor for observations in groundwater | Corrective factor for observations in groundwater | Yes then CFgw = 1<br>No then CFgw = 1 | |

**Questions:**
Do you agree on the proposed revision of the Exposure score for Categories 1,3 and 6?

Do you agree on the proposed weights and aggravating factors?

### *7.1.2 Revising the "Exposure" score for categories 2,4 and 5*

*Current status*
The current exposure score for Categories 2, 4 and 5 is a "Predicted Exposure" score derived as the sum of tonnage score + use pattern score (see ref. NORMAN Prioritisation scheme).

*Critical issues*
Current experience with prioritisation studies in Europe has shown that for certain use categories information on tonnage is not accessible (confidential data). For a significant number of substances, it would be necessary to apply default values but this is not satisfactory for an indicator / score which is expected to discriminate the compounds. Moreover, tonnage and use pattern alone are not satisfactory indicators to predict environmental exposure.

Today, data from NTS and literature databases are not yet used as potential sources of information in the calculation of the Exposure score. These new sources of information, provided that the data are adequately stored in databases, could well be systematically used to derive a more robust exposure score.

*Proposals*
We propose the definition of a new Exposure score for Cat. 2, 4 and 5, made of 3 sub-scores:

1. "NTS data" score
2. "Literature data" score
3. "Exposure index" score

These scores would then be aggregated with different coefficients based on the quality, the uncertainty and availability of these data, as described below (see 1 to 4 and final formula for the calculation of the EXPO score).

1. NTS data score

| Categories | Indicators | Scoring rules |
|---|---|---|
| Cat. 2, 5 (all sub-categories) | Nb. countries with positive detections (or x%) | (0-1) |
| | Frequency of positive detections | FoD (0-1) |
| | Extent of Exceedance = MEC95/Lowest PNEC | EoE (0-1) |
| | Frequency of Exceedance = n / N | FoE (0-1) |
| Cat. 4A, 4B | Nb. countries (or x%) where the component is detected | (0-1) |
| | Frequency of Appearance of component | FoA (0-1) |
| Cat. 4F | Nb. countries (or x%) where the m/z hit is detected | (0-1) |
| | Frequency of Appearance | FoA (0-1) |

NOTE:

The difference between the indicators for Cat. 2 and 5 and those proposed for Cat. 4 is explained by the fact that:

- Cat. 2 and 5 address only peaks which have been identified with a certain confidence (semi-quantification is possible);
- Cat. 4A and B address peaks for which identification is not proven with sufficient confidence (qualitative assessment);
- Cat. 4F could address m/z which are occurring with a certain frequency in a certain number of countries for identification.

The weight of the different factors for the derivation of the “NTS score” will be defined in order to have a final score comprised between 0 and 1, regardless of the category.

**Question**
Do you agree with the proposed ranking indicators?

2. “Literature data” score

| Categories | Indicators | Scoring scale |
|---|---|---|
| Cat. 2, 5 | Nb of regions with quantified data (or%) | |
| | Number of publications with positive detections | |
| | Number of positive detections (or%) | |
| | Number of matrices with positive detections | |
| | Nb of publications with exceedance of PNEC | |
| Cat. 4 | To be proposed | |
| Cat. 4F | Not applicable (value by default?) | |

The equation and the weight of the different factors for the derivation of the “Literature score” will be defined in order to have a final score between 0 and 1, regardless of the category.

**Question**
Do you agree with the proposed ranking indicators? They are applicable to all compounds that can be identified by a name and a CAS N° / structure.

In Cat 4 by definition we do not have positive detections (LOQ > PNEC and compounds identification is not proven with sufficient confidence in NTS data). What alternative indicators could be proposed?

Cat 4F: the “literature data” score is not applicable because in this sub-category we have m/z hits without a name, or structure. Do we need to apply a default value?

3. Exposure index

The Exposure Index scores developed by KEMI for industrial chemicals and by Fh-IME for biocides (see document enclosed) have already been discussed within WG-1 as potential replacement of the current “Predicted Exposure” score. The background documents explaining in detail the concept of these two indices are available for comments.

They are based on three different indicators: the volume (or a surrogate of the volume), the potential for release during use and the widespread use of the compound.

| Categories | Indicators | Scoring scale |
|---|---|---|
| Cat 2, 4, 5 | Annual tonnage (or a surrogate of annual tonnage) | (0-1) |
| | Release during use | (0-1) |
| | Wide dispersive use | (0-1) |
| Cat 4F | Not applicable (value by default?) | |

The same indicators can be applied to all categories, except for Cat 4F.

The final Exposure Index is a value comprised between 0 and 1, resulting from the equations proposed by KEMI and by Fh-IME.

When applying this Exposure Index it has, however, to be noted that for each of these components it has been necessary to develop a relevant proxy, depending on the type of information available for each category of compound. For example the exposure index developed by Fh-IME for biocides makes use of qualitative information on the amount / volume used for the different biocidal active substances to calculate a proxy for the missing tonnage data.

These Exposure Indices represent the state of the art in prioritising compounds for which monitoring data are missing. Various improvements have already been made, but some research is necessary to keep improving these indices. First, they do not consider physico-chemical properties, thereby limiting their predictive power. Second, they are not harmonised so as to allow comparison of scores derived from different indices. For example, 0.6 calculated with the Fh-IME biocide index cannot be compared to a score of 0.6 calculated with the KEMI index. For future improvements, statistical developments should be envisaged.

**Questions:**

Do we have an exposure index for each compound in SusDat?

For biocides we may have two exposure indices, one Exposure Index provided by KEMI for their use as industrial chemicals and another for their specific use as biocides (Fh-IME methodology). What is the value that we need to use for ranking?

When we have no Exposure Index available, what default value should be used?

4. Quality score

We propose to develop a quality score to qualify the different sub-scores and adjust their weights.

The NTS score could be based on "Identification Proofs", the Literature score could be based on Number and quality of publications, and the Exposure Index on uncertainty grades (e.g. as developed by KEMI).

For the NTS score, the different grades of the "Quality score" could be defined on the basis of the following considerations / parameters:

- Known experimental fragmentation behaviour or predicted behaviour
- In the case of known experimental fragmentation behaviour, the minimum number of fragments can be specified as a function of the molecular weight. Numbers should be low (1-2), because there is enough confidence coming from the fact that the compound is present in spectral libraries.
- In the case of predicted fragmentation behaviour, the minimum number of fragments should be specified as a function of the molecular weight (Big molecule = big puzzle; Small molecule = small puzzle (puzzle for kids); Number of fragments = number of puzzle pieces that match.
- Presence of fragments of diagnostic importance (e.g. 79.9574 is diagnostic fragment of $SO_3$ group in negative ionisation).
- Expert judgement

Here we provide an example of possible criteria for discussion (NOTE: the number of quality grades, 5 in this example, can be reduced):

| **Quality grade** | **NTS score** | **Literature data** | **Exposure index** |
|---|---|---|---|
| 5 (highest quality) | To be discussed (see proposals above) | More than X "good" pub. | 1-to-5 scale, as defined by KEMI[5] |
| 4 | | More than Y "good" pub. | |
| 3 | | 1 "good" pub. | |
| 2 | | Only questionable & unknown pub. | |
| 1 (lowest quality) | | Only questionable pub. | |

[5] I am aware that the score by KEMI is defined as "1" for highest quality (minimun uncertainty) and "5" for lowest quality (max uncertainty) but this is just a proposal and if we want to apply a quality score we can then decided how to play with these existing values.

Final formula : $EXPO\_cat2,4,5 = \frac{NTS_{quality}*NTS_{score}+Lit_{quality}*Lit_{score}+EI_{quality}*EI_{score}}{NTS_{quality}+Lit_{quality}+EI_{quality}}$

**Questions**

Do you agree with the application of a quality score? Or should we simply add the three sub-scores to obtain the final EXPO score?

## 7.2 Hazard Score

### *7.2.1 A multi-criteria approach for calculation of the hazard score (based on PNEC and CMR)*

*Current status*
The current "Hazard" score is calculated as the average of the following indicators (see table below).

| Score | Indicator | Description | Formula |
|---|---|---|---|
| Environmental Hazards | H) PBT /vPvB | Overall PBT/vPvB score = [(P + B + T) individual scores + (PBT/vPvB) score] / 4 | See NORMAN Prioritisation Framework (Table 9 and 10 in Annex II) |
| | I) LRAT (long-range air transport) | Half-life ($t_{1/2}$) in air >2 days and Vapour Pressure (VP) < 1000 Pa | $t_{1/2}$ in air >2 days and VP <1000 Pa = 1<br>$t_{1/2}$in air ≤2 days and /or VP ≥ 1000 Pa = 0 |
| | J)Non-standard endpoints | Examples:<br><br>hatch size | Non standard endpoints present = 1<br>Under examination = 0.5<br>Not examined = 0.25<br>Evaluated and classified not toxic = 0 |
| Human Health Hazards | K) CMR = Max («Carcinogenicity», «Mutagenicity», «Reprotoxicity ») | The CMR final score is then derived as the highest value between the individual carcinogenicity, mutagenicity and reprotoxicity scores. | CMR, category 1 = 1<br>CMR, category 2 = 0.75<br>CMR, categorie 3 = 0.5<br>Under examination = 0.5<br>Examined and info not suff. = 0.25<br>Not examined = 0.25<br>Examined and not classified = 0 |
| | L) Endocrine disrupting properties | | Proven ED = 1<br>Suspect ED = 0.5<br>Not examined = 0.25<br>Not proven ED = 0 |

*Issues*

- PNEC are included in the calculation of the PBT score. However, with the currently used algorithm, PNEC account for only 5% of the overall Hazard score[6].

- Similarly, CMR properties are "diluted" in the Hazard score and account for only 20%, whereas they should be the main contributor to this score along with PNEC. This is due to the proliferation of factors directly included in the derivation of the "Hazard" score.

- Besides PBT / vPvB substances, new indicators should be introduced to prioritise PM(T) / vPvM substances. To be discussed based on input from on-going projects (PROMOTE projects, etc.)

*Proposal*
The hazard score should be based on two main components: a first one reflecting ecotoxicological hazard (Lowest PNEC), a second reflecting human toxicity (CMR properties).

The proposed scale for the "ecotoxicological hazard" score (Ecotox score based on Lowest PNEC) is illustrated in the table below.

**Table 6: Formula for Ecotox score derivation**

| Criteria | Score |
|---|---|
| PNEC | Score between 0 and 1:<br>≤0.1 µg/l : 1<br>≤1 µg/l : 0.75<br>≤10 µg/l : 0.5<br>≤100 µg/l : 0.25<br>>100 µg/l : 0<br>Default value (when data not available) = 0.25 |

The scale for the "human toxicity" score (based on CMR properties) should remain as it is.

Moreover, we suggest the introduction of an aggravating factor to be added to the Ecotox and human toxicity scores.

We propose the following formula for the aggravating factor (AF) based on PBT/PE properties:

- *If Substance = PBT or vPvB OR Substance = ED (proved), then AF= 0.2*
- *Else if Substance = PE (presumed) then AF=0.1*
- *Else AF=0*

The final formula for "Hazard" is reported in the table below.

Summary Table:

[6] PBT/vPvB score is calculated as [(P + B + T) individual scores + (PBT/vPvB) score] / 4. PNEC is considered in the derivation of T. This PBT score is then divided by 5 in the calculation of the overall hazard score. Hence, the T factor account for 1/20 of the overall Hazard score.

| Hazard sub-scores | Selected criteria | Score formula | Final score |
|---|---|---|---|
| Ecotoxicity score "Ecotox" | PNEC | Score between 0 and 1:<br>≤0,1 µg/l : 1<br>≤1 µg/l : 0.75<br>≤10 µg/l : 0.5<br>≤100 µg/l : 0.25<br>>100 µg/l : 0<br>Default value (when data not available) = 0.25 | HAZ = ½ (Ecotox+Tox) + AF<br><br>IF HAZ>1, HAZ=1 |
| Human toxicity score "Tox" | CMR | CMR, category 1 = 1<br>CMR, category 2 = 0.75<br>CMR, categorie 3 = 0.5<br>Under examination = 0.5<br>Examined and info not suff. = 0.25<br>Not examined = 0.25<br>Examined and not classified = 0 | |
| Aggravating factor "AF" | PBT/ED | If Substance = PBT or Substance = PE (proved) Then AF= 0.2<br>Else if Substance =ED (presumed) Then AF=0.1<br>Else AF=0 | |

Question:
Do you agree on the proposed weights and aggravating factors?
Should we discard the LRAT indicator as proposed?

### 7.3 How to justify scores associated with default values?

To be discussed?

## 8 New indicators to improve prioritisation of compounds contributing to mixture effects (potential frequent contributors)

To be drafted

…

| Categories / indicators | Cat. 1 | | Cat. 2 | | | Cat. 3 | Cat. 4 | | | Cat. 5 | | | Cat. 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1A | 1B | 2A | 2B | 2F | 3 | 4A | 4B | 4F | 5A | 5B | 5F | ~~6A~~ | ~~6B~~ 6 |
| Analyses available in relevant matrix(ces) | Yes | Yes | Yes/No | Yes | No data | Yes | Yes or No data | Yes | Yes OR No data | Yes/No | Yes | No data | ~~Yes~~ | Yes |
| **≥ 4 countries with analysis** | Yes | Yes | <4 countries *AND/OR* <100 sites | Yes | No data | Yes | <4 countries *AND/OR* <100 sites | Yes | - | <4 countries *AND/OR* <100 sites | Yes | No data | ~~Yes~~ | Yes |
| **≥100 sites with analysis** | Yes | Yes | | Yes | No data | Yes | | Yes | - | | Yes | No data | ~~Yes~~ | Yes |
| ~~≥ 20~~ 50 sites with analyses > LOQ (recent data) | Yes | No | - | No | No data | Yes OR | - | No | - | - | No | No data | ~~Yes~~ | - |
| ~~LOQmax <PNEC~~ ≥ 100 sites with LOQmin < PNEC | - | Yes | - | No | No data | No AND $LOQ_{max}$< PNEC | No or No data | No | No or No data | - | No | No data | - | Yes |
| **LOQmin <PNEC** | - | Yes | $LOQ_{min}$ (datasets) < PNEC *OR* $LOQ_{literat}$ <PNEC | $LOQ_{min}$ (datasets) < PNEC *OR* $LOQ_{literat}$ <PNEC | No data | - | No or No data | No | No or No data | $LOQ_{min}$ (datasets) < PNEC *OR* $LOQ_{literat}$ <PNEC | $LOQ_{min}$ (datasets) < PNEC *OR* $LOQ_{literat}$ <PNEC | No data | - | Yes |
| **$LOQ_{literat}$ <PNEC** | - | - | | | Yes | - | No | No | No data | | | Yes | - | - |
| **Suff. data for hazard assessment** | Yes | Yes | Yes | Yes | Yes | No | - | - | - | No | No | No | ~~Yes~~ | Yes |
| **Potential risk identified ($MEC_{95}$ /Lowest PNEC≥1)** | Yes | No | - | - | No data | - | - | - | No data | - | - | No data | ~~No~~ | No |
| Potential risk identified MEC25th_site / PNEC> 1 | Yes | Yes | - | - | No data | - | - | - | No data | - | - | No data | | No |

Table X: **NORMAN framework** with proposed changes for sufficiently monitored compounds - List of indicators and cut-off values applied for the allocation of the candidate substances to action Categories 1 to 6

# References


1. Dulio V, von der Ohe PC: **NORMAN Prioritisation framework for emerging substances**: ISBN: 978-2-9545254-0-2; 2013.
2. Dulio V, von der Ohe PC, Slobodnik J, NORMAN Prioritisation WG: **The NORMAN approach for setting priorities among emerging contaminants in Europe: Working Group 1**; 2011.
3. Dulio V, Andrès S: **Référentiel méthodologique pour la priorisation des micropolluants des milieux aquatiques établi par le Comité d'Experts National pour la priorisation des micropolluants aquatiques (CEP).** . *Programme scientifique et technique 2012*.
4. Botta F, Dulio V, Andres S, Feray C, Morin A: **A watch list of emerging pollutants for surface water monitoring in France**. *NORMAN Network Bulletin* 2012, **3**:7-8.
5. **NORMAN SusDat: Suspect List Exchange** http://www.norman-network.com/datatable/
6. **EMPODAT: Database of geo-referenced monitoring data on emerging substances in air, water and soil [**http://www.normandata.eu/empodat_index.php?menu_type=2
7. NORMAN Network: **Norman MassBank. [**http://massbank.ufz.de/MassBank/**] []**. In.


To be completed