European Environment Agency
European Topic Centre
Biodiversity and ecosystems

Methodology, oxygen consuming substances and nutrients indicators

(WAT002-WAT003)

Extended supporting information for the published indicators

Version: 3
Date: 23.09.2025
ETC BE task number: 1.3.45.1

Author: Kari Austnes
From: NIVA

Contributors: Jan-Erik Thrane
From: NIVA

Document History

Version	Date	Author(s)	Remarks
1.0	12.12.2024	ETC BE	First draft for consultation on new methodology
2.0	27.06.2025	ETC BE	Update for the consultation on indicators 2025
3.0	23.09.2025	ETC BE	Minor updates for publication

1 Introduction

The objective of this document is to describe all steps and procedures involved in the preparation of the data and plots included in the freshwater indicators Oxygen consuming substances in European rivers (WAT002) and Nutrients in freshwater in Europe (WAT003). An overview of the methodology is provided in the Supporting information of the indicators, but that section does not allow the level of detailed technical information provided here. The document will be updated every year, at the end of the indicator production process, to reflect any changes made to the procedure.

The indicators have undergone some changes over the years, in particular changing to a more condensed format in 2022. However, the data outputs have been the same:

- Aggregated time series for Europe: Averaging complete time series. Gap filling of gaps up to three years has been allowed, to increase the number of complete series.
- Trend analysis: Mann-Kendall trend analysis of the same time series as used in the time series plots, excluding gap-filled values.
- Present state analysis: Distribution of monitoring sites/water bodies to concentration classes per country, based on average concentrations for the last three years and fixed concentration thresholds.

In 2025 the indicator monitoring was reviewed and a new methodology suggested. After some adjustment, this methodology was used for the first time to produce the indicators in 2026. This methodology is what is described in this document. The main focus of the review was to increase spatial representativity of the time series. There was also a wish to change from analysis at monitoring site to water body (WB) level. In the new methodology site time series are aggregated to WB time series and gap filled using GAM modelling. The WB time series are averaged to country time series, which are weighted by water size in the subsequent average to European level. Altogether this allows more site time series to be included while also avoiding a skewed picture by giving too much weight to countries with a high monitoring density. There were also other minor changes, including introducing confidence intervals in the time series plots and basing the present state analysis on quintiles and data from the most recent six years. The procedure for producing annual data per site remains the same, although minor adjustments are made every year.

2 Input data

2.1 WISE SOE WISE-6 data

The indicators build on data extracted from Waterbase – Water Quality ICM (WISE-6). This is one of the WISE State of Environment (SoE) databases and is administered by the European Environment Agency (EEA). The reporting obligation for Waterbase – Water Quality ICM is an EIONET core data flow. This means that the data represent the monitoring network of Member States of the EU as well as other EIONET reporting countries. Data are reported annually under Reportnet3. Reporting obligation, guidelines, data dictionary and other information is found on the Eionet Central Data Repository page for WISE-6.

The WISE-6 database contains water quality data on a range of determinands in inland (rivers, lakes and groundwater), coastal and marine waters. The data can be downloaded from EEA web pages or EEA geospatial catalogue, or extracted in Discodata (WISE SOE WISE-6 tables).

Data can be reported as disaggregated, aggregated or for selected groundwater determinands (including groundwater nitrate, used in WAT003) as aggregated to water body (WB) level. Disaggregated data are sample data, where a record represents a specific monitoring site, date, determinand, matrix and depth. Aggregated data combines sample data to one annual record for the given monitoring site, determinand, matrix and depth. Data aggregated to WB level combines annually aggregated data for different monitoring sites belonging to the same WB. Rules for aggregating data are provided in the data dictionary. As of 2015, reporting disaggregated records is encouraged, while before this it was not possible to report disaggregated data for the WAT002 and WAT003 determinands. Data from before 2013[1] is only available as aggregated data unless data has been resubmitted after 2015. The different types of data are available in three separate tables, here listed as they can be downloaded from the datahub:

Waterbase_T_WISE6_DisaggregatedData

Waterbase_T_WISE6_AggregatedData

Waterbase_T_WISE6_AggregatedDataByWaterBody

2.2 WISE Statistics

In the WISE Statistics (called WISE Indicators in Discodata) tables, data have been further processed for the purpose of indicators and other products. In the [AggregatedData] table, disaggregated data and annually aggregated data are combined into annual data. Here records with certain QC statements are excluded. Other records are also excluded according to pre-defined principles. BOD7 and total oxidised nitrogen data are used to gap fill BOD5 and nitrate data, respectively, and some corrections are made. Disaggregated data are aggregated to annual means, following the principles outlined in the data dictionary as to how to e.g. treat values below LOQ. Duplicates are excluded following certain rules. The resulting tables are sets of quality controlled, annual mean values per site for selected determinands. Further details on the WISE Statistics [AggregatedData] table is provided in chapter 3.

In the WISE Statistics table [AggregatedDataByWaterBody] annual site data are aggregated to water body (WB) data by average across the sites available per year. Data reported as aggregated to WB (groundwater nitrate) are included when reported disaggregated or aggregated data are not available. Further details on the [AggregatedDataByWaterBody] table are provided in chapter 7.1.

Note that [AggregatedData] and [AggregatedDataByWaterBody] contains gap-filled values according to the old methodology. These are not used with the current methodology. Only records with qcGapFilling = 'Using reported data.' is used.

In the production of the indicators, WISE Statistics data are extracted directly from EEA's Common WorkSpace (CWS), but the same tables are available in Discodata, listed under [WISE_Indicators].[latest]. Further background is available in the EEA geospatial data catalogue.

[1] The latest data reported are from the year prior to the reporting year. There was no reporting in 2014, so in 2015 both 2013 and 2014 data were reported.

2.3 Quality control

There is a range of quality control steps involved, both in production of the WISE-6 and the WISE Statistics tables:

1) Automatic tests are run upon data submission in Reportnet3. The latest version of the tests is available here. Only BLOCKER messages prevent data submission. However, also ERROR and WARNING messages should be carefully examined.
2) Further tests and checks of recently submitted data are run by EEA, providing the final feedback to the countries. Clarifications and/or corrections may be asked for.
3) When the new data are incorporated in the database, automatic tests are run again. This includes z score outlier tests of the disaggregated data for each determinand, giving the following metadata statement labels:
   - QC_OUTLIER_STDEV: Checks values against the other values from the same site. Flagged if z score > 5.5
   - QC_OUTLIER_STDEV_YEAR: Checks values against the other values from the same site and year. Flagged if z score > 3
4) After a first run of WISE Statistics, further quality control steps are made
   - The WISE Statistics processing includes additional z-score tests.[2] Time series where one or more values are flagged based on these tests are plotted and examined to check if there are reasons to flag records causing the deviations, going back to the initial disaggregated records where available. Only extreme records are flagged permanently (receiving metadata statements in the WISE-6 database, starting with QC_OUTLIER_EXPERT). The tests are run
     - per time series: z-score cutoff = 4
     - per determinand/water category combination across all years: z-score cutoff = 5.5
     - per determinand/water category combination in a specific year: z-score cutoff = 5.5
   - The WISE Statistics and preliminary indicator output is closely examined, looking for deviations in e.g. aggregated country time series or data standing out in visual inspection of the tables. This can identify issues that may have passed automatic tests, e.g.:
     - step changes
     - when all data for one year or several years from a country deviate from the remaining years
     - when there are two or more suspicious values in a time series
     - unit errors
     - values resulting from very high LOQ

     Again, only extreme records are flagged in the WISE-6 database

If [resultObservationStatus] = 'A', automatic outlier flags are not added. This is why it is important to report this correctly, as stated in the Data Dictionary: “Use the 'A' observation status flag to confirm that the given record is correct. The flag should not be applied to all valid records. It is intended for confirmation of extremely high or low values or for other special cases

[2] These are run on the combined set of calculated annual and reported annual data, i.e. before removal of duplicates (see section 3.5 and 7.1)

where confirmation is needed or is relevant." Manual flags are generally not added to records with [resultObservationStatus] = 'A', but rare exceptions occur when e.g. when [resultObservationStatus] = 'A' is generally applied and not only to single records.

After the quality control based on the first WISE Statistics run, the WISE-6 database is updated with the additional metadata statements and finalised. Then WISE Statistics is run again, and the indicators are produced.

The metadata statements can be found in the WISE SOE WISE-6 tables. The countries should check these and re-submit data correct data as far as possible. If records have outlier statements, but the values can be confirmed as not being outliers, the [resultObservationStatus] = 'A' can be applied. Other issues of concern may be reported directly to the countries, like missing data, issues with site or WB codes, possible unit or decimal errors, suspiciously high LOQ etc. Addressing these issues will improve the quality and coverage of the indicators.

2.4 Spatial data

With the new indicator methodology, the total size of all WBs for each country is needed, i.e. total length for rivers and total area for lakes and groundwater (see chapter 5). This is collected from the WISE Spatial database, selecting data reported under the Water Framework Directive (WFD) for countries reporting there, otherwise data reported under Eionet. For WFD, data reported under the 2nd RBMPs is selected for countries not yet reporting under the 3rd RBMPs.

Countries not reporting under the WFD report spatial data only for WBs for which they report data. This means that the WBs may not cover all WBs in the country. However, inspection of coverage for countries with WISE-6 data that can be included in the time series analyses, indicates that the coverage is sufficient to be representative for the country. Countries not reporting spatial data or reporting incorrect spatial data must be excluded from the time series analysis. In practice there are few such exclusions.

3 Annual data per monitoring site

Calculating annual data per monitoring site and combining these with data reported as annual data per site is done in the automatic WISE Statistics processing. The results are available in the [AggregatedData] table. The details of this processing are described in this chapter.

3.1 Selection of data

The first step in the WISE Statistics processing is to select data to go into the processing. The input data for WAT002 and WAT003 is the determinand and water category combinations given in Table 1. In practice more data is included in the data processing, as given in the table [AncillaryData_TimeSeries] in WISE Statistics (WISE_Indicators in Discodata).

Table 1. Determinand and water category combinations used for the WAT002 and WAT003 indicators. Column headings (except first and last) reflect those used in the WISE_SOE tables.

Indi-cator	observedProperty DeterminandCode	observedProperty DeterminandLabel	resultUoM	parameter WaterBody Category	Comments
WAT 002	EEA_3133-01-5	BOD5	mg{O2}/L	RW	
	EEA_3133-02-6	BOD7	mg{O2}/L	RW	Converted to BOD5
	CAS_14798-03-9	Ammonium	mg{NH4}/L	RW	Converted to µg NH_4-N/L
WAT 003	CAS_14797-55-8	Nitrate	mg{NO3}/L	GW	
	EEA_3161-02-2	Total oxidised nitrogen	mg{N}/L	GW	Assumed to be equivalent to nitrate, converted to mg NO3/L
	CAS_14797-55-8	Nitrate	mg{NO3}/L	RW	Converted to mg NO3-N/L
	EEA_3161-02-2	Total oxidised nitrogen	mg{N}/L	RW	Assumed to be equivalent to nitrate
	CAS_14265-44-2	Phosphate	mg{P}/L	RW	
	CAS_7723-14-0	Total phosphorus	mg{P}/L	LW	

The following inclusion criteria are used:

- [observedPropertyDeterminandCode] and [waterBodyCategory] combinations in the table [WISE_Statistics].[AncillaryData_TimeSeries] where the column [flag] is not empty
- [metadata_statusCode] in the WISE_SOE WISE-6 tables (see chapter 2.1) is 'accepted', 'valid', 'experimental', 'stable' or 'derived'
- data between 1989 and two years prior to the indicator production year (only data from 1992 onwards are currently used for the indicators)
- [procedureAnalysedMatrix] is 'W', 'W-DIS' or 'W-SPM'

The following exclusion criteria are used:
- [resultObservationStatus] is 'L', 'M', 'N' or 'O', meaning the value is missing, or 'Z', meaning that the record is marked for deletion

The automatic and manual quality control (see chapter 2.3) leads to certain records receiving a metadata statement. Some of these metadata statements lead to exclusion from the indicators.

The various statement labels and whether they cause exclusion are given in the table which in Discodata is called [WISE_Indicators].[latest].[AncillaryTable_QualityControl][3]. This includes:

- Records with values above or below the extreme limits for a given substance
- Records that fail the standard deviation tests run on disaggregated data (point 3 in chapter 2.3)
- Data with unknown monitoring site location
- Data with unit issues
- Records failing the expert QC
- Records with LOQ issues (see ch. 3.3)

3.2 Conversions

To increase the amount of data available, BOD7 is converted to BOD5 and combined with BOD5 to give the indicator determinand BOD5, and total oxidised nitrogen is combined with nitrate to give the indicator determinand nitrate. Some unit conversions are made to produce the desired indicator units, see [WISE_Statistics].[AncillaryData_TimeSeries]. For nitrate mg NO_3/l is commonly used for groundwater, while mg NO_3-N/l (mg N/l) is more common for surface waters, so these are the units that have been chosen.

3.3 Handling of limit of quantification and corrections

Countries have been encouraged to report LOQ along with each data record since 2010 and have been required to do so for data reported since 2015. Actual LOQ is requested for disaggregated data. For aggregated data, LOQ is defined as described in section 3.4. Countries also report whether the reported value is below LOQ.

Table 2 summarises how different combinations of reported value and LOQ value information are handled. Cases 1-3 and 8 are correct reporting, case 5 is ambiguous while the remaining are incorrect or highly suspicious. Cases 6-7, 9, 11 and 13 are considered so serious that they cause exclusion of records. Case 4 is incorrect and has been considered for exclusion. Cases 10, 12 and 14 are considered nonessential errors.

Setting observed values reported as being below LOQ to half LOQ before calculating the mean value (case 1-4) is in accordance with the QA/QC Directive (Commission directive 2009/90/EC). If the reported value is missing for values reported as being below LOQ (as is in accordance with the Data Dictionary for disaggregated data) it is set equal to half LOQ for observed values (disaggregated data) or LOQ for mean values (aggregated data), i.e. in line with case 1-3.

Table 2. Overview of LOQ cases and how they are handled. Reported value refers to observed value (disaggregated data) or mean value (aggregated data), if not otherwise stated in footnote

	IF	THEN	Example LOQ	Example value	Example OUTPUT	statementLabel

[3] Only the statement label is given here. The more detailed statement messages are available in the WISE SOE WISE-6 tables

1	reported value = reported LOQ value and flagged as below LOQ	reported LOQ value/2[4]	0.5	0.5	0.25	
2	reported value = reported LOQ value/2 and flagged as below LOQ	reported LOQ value/2[5]	0.5	0.25	0.25	
3	reported value < reported LOQ value (but not half of it) and flagged as below LOQ	reported LOQ value/2[5]	0.5	0.1	0.25	
4	reported value > reported LOQ value and flagged as below LOQ	reported LOQ value/2[5]	0.5	3	0.25	QC_BELOW_LOQ_TRUE_OBSERVED (QC_BELOW_LOQ_TRUE_MEAN for aggregated data)
5	reported value = reported LOQ and not flagged as below LOQ	reported value	0.5	0.5	0.5	
6	reported value = reported LOQ value/2 and not flagged as below LOQ	exclude[6]	0.5	0.25	exclude	QC_BELOW_LOQ_FALSE_OBSERVED (QC_BELOW_LOQ_FALSE_MEAN for aggregated data)
7	reported value < reported LOQ value (but not half of it) and not flagged as below LOQ	exclude[6]	0.5	0.1	exclude	QC_BELOW_LOQ_FALSE_OBSERVED (QC_BELOW_LOQ_FALSE_MEAN for aggregated data)
8	reported value > reported LOQ value and not flagged as below LOQ	reported value	0.5	3	3	
9	no reported LOQ value available, but reported value flagged as below LOQ	exclude	not reported	0.1	exclude	QC_LOQ_UNKNOWN_BELOW (QC_LOQ_UNKNOWN_MEAN_BELOW_LOQ for aggregated)
10	no reported LOQ value available and reported value not flagged as below LOQ	reported value	not reported	0.5	0.5	QC_LOQ_UNKNOWN
11	reported LOQ is <=0, but reported value flagged as below LOQ	exclude	<=0	0.5	exclude	QC_LOQ_BELOW_MIN (QC_LOQ_BELOW_MIN_MEAN_BELOW_LOQ for aggregated)
12	reported LOQ is <=0 and reported value not flagged as below LOQ	reported value	<=0	0.5	0.5	QC_LOQ_BELOW_MIN_NONESSENTIAL
13	reported LOQ is > max LOQ, but reported value flagged as below LOQ	exclude[7]	5	0.5	exclude	QC_LOQ_ABOVE_MAX
14	reported LOQ is > max LOQ and reported value not flagged as below LOQ	reported value	5	0.5	0.5	QC_LOQ_ABOVE_MAX_NONESSENTIAL

In addition some corrections to the data are made:

Disaggregated data:

[4] Not for reported aggregated data – here the reported mean is used
[5] For reported aggregated data the mean is set to the LOQ value (not legacy data)
[6] Not for reported aggregated – here the reported mean is used
[7] Not for reported aggregated – here the reported mean is used

- For determinands where LOQ is inapplicable, LOQ is set to missing where reported and the sample value is set to not being below LOQ if it was reported as being below LOQ

Reported aggregated data by site:

- If any of the summary values are reported as being below LOQ, but the number of samples below LOQ is zero, the number of samples below LOQ is set to missing
- For determinands where LOQ is inapplicable, LOQ and number of samples below LOQ are set to missing where reported and the summary values are set to not being below LOQ if it was reported as being below
- For legacy data[8] where the mean is reported as being below LOQ or there is no information whether it is below LOQ, LOQ and number of samples below LOQ are set to missing and the summary values are set to not being below LOQ. The exception is where the mean value is missing or equal to LOQ or LOQ is missing. This is to avoid the reported mean value to be replaced by the LOQ when reported as being below LOQ, which is the normal procedure (see case 2-3 in Table 2). For legacy data it is observed that the LOQ reporting is frequently suspicious, while the mean values seem plausible
- When the sample depth is negative it is set to missing

3.4 Aggregation

Aggregation to annual values per monitoring site

The aggregation to annual values is done in accordance with the description in the Data Dictionary. Summary values are calculated, i.e. mean, LOQ, LOQ flags (whether value is below LOQ), number of samples, number of samples below LOQ:

- The mean value is calculated across all sample values from the same monitoring site and year, i.e. across dates, matrices (see section 3.1), determinands (where more than one, see section 3.2) and depth. Only a small fraction of the data is reported for multiple matrices and depths on the same date, though. If the sample value is below LOQ, LOQ/2 is used as sample value in calculation of the mean (Table 2).
- If any of the sample values are reported as being below LOQ, the aggregated LOQ is set to the highest LOQ among these samples. Otherwise, the aggregated LOQ is set to the highest of all the sample LOQs.
- Whether the mean is below LOQ is evaluated against the aggregated LOQ. If the mean is below the aggregated LOQ it is set equal to the aggregated LOQ. There is an exception where all sample values are above LOQ, while the mean is below the aggregated LOQ. Then the original mean value is kept and the mean is set as not being below LOQ.
- Number of samples is counted as the number of disaggregated records per year.
- Number of samples below LOQ is counted as the number of disaggregated records where the sample value is reported as being below LOQ.

3.5 Duplicates

Wherever there are several annual records for the same (converted) determinand, water category and year, duplicates are selected according to the following rules:

1) Records which pass the WISE Statistics z score tests are given priority
2) Source of data is given priority in this order:
    a. Annual data derived from sample data (disaggregated data)

[8] Data reported under the old reporting system, i.e. data from 2012 and backwards that have not been resubmitted after December 2015

   b. Reported annual data (aggregated data)
   c. Reported annual data for alternative determinands BOD7 or total oxidised nitrogen
   d. Annual data that are legacy data or modified records
   e. Annual data for alternative determinands that are legacy data or modified records
3) For fractions data is given priority in the following order:
   a. Total
   b. Dissolved
   c. Combination of total and dissolved within the same year
   d. Suspended particulate matter (spm) or a combination of spm and other fractions within the same year
4) Records with highest number of samples is given priority
5) Records with lowest LOQ is given priority
6) Records where mean value is not missing
7) Records with lowest sample depth (closest to zero) is given priority

4 Annual data per water body and gap filling of time series

4.1 Analysis at water body level

Up until 2024, data aggregated to monitoring site level was used as input to the indicators, except for groundwater nitrate, where water body (WB) data has been used.

Changing to analysis at WB level has several advantages:

- Data from sites within the same WB are not completely independent. It is more statistically sound to handle sites from the same WB differently from sites from different WBs
- Combining site time series for different sites to WB time series may fill gaps in the time series and thereby provide more time series that qualify for inclusion in the time series analysis (see chapter 4.2)
- Aggregating to WB level removes some of the bias towards countries with dense monitoring networks, i.e. in the cases where the number of monitoring sites per WB is higher than in other countries
- There is a finite number of WBs, which can be used in analyses
- With WBs as analysing unit it is possible to link in WFD information if needed
- Groundwater nitrate has always been calculated at WB level (due to few long time series at site level and some data reported directly at WB level), so using WB level for all determinands gives a more harmonised approach

There are, however, some disadvantages as well:

- Uncertainty is introduced by combining site time series of different length to WB time series
- Not all sites are linked to a WB, in particular for non-EU countries
- The definition of WBs is heterogeneous

- The link between monitoring site and WISE-6 WBs is not always up to date with the most recent WB delineation, meaning that the WISE-6 WBs are not a true subset of the countries' total set of WBs

The current methodology reduces the impact of most of these disadvantages. However, the lack of link between monitoring sites and WBs remains an issue in some cases. All groundwater sites with nitrate data can be linked to a WB, but for surface waters there is not always such a link. This goes especially for non-EU countries, but there are also sites from EU countries without WB information. The latter are largely sites with old data only and/or they are marked for deletion (wiseEvolutionType = 'deletion'). For the time series analysis, monitoring sites without WB information must be excluded. In practice, this only affects a couple of countries, given the availability of data. In the present state analysis, monitoring sites without WB information have been included for certain countries (see chapter 7).

4.2 Data selection

For time series analysis, a time range must be defined. This should largely be guided by data availability. Up until 2024, the start years 1992 and 2000 have been used, with the end year being the most recently reported data (two years before the indicator publication year). Using 1992 as start year keeps the long-term perspective, showing the large changes happening since before e.g. the WFD was implemented. However, as several countries started reporting later than this, an additional time range with a later start year covers more countries and WBs. On the other hand, a very short time range provides little information about change. The start year 2000 was selected for the shorter time range many years ago. Analysis has showed that there was a marked increase in the number of WBs with relevant data around 2007. Hence, the start year for the shorter time series was changed from 2000 to 2007 with the indicators produced in 2025. This makes it possible to include far more WBs, which increases the spatial representativity while the time range is still sufficiently long to detect changes.

When combining WB time series to country or European time series, it is essential that the time series are complete, i.e. that there is data from each year in the time range. This is to avoid changes in concentration over time that may be interpreted as real changes, while in practice they only reflect differences in number of WBs with data between years. However, as few time series are complete, gap filling is allowed, to increase the number of available time series and thus spatial representativity.

Up until 2024, gaps up to three years were allowed, but with the current, more robust, methodology for gap filling (see next chapter), longer gaps can be allowed. Setting the criteria for data requirements is a trade-off between uncertainty introduced by the gap filling and spatial representativity. This has been tested thoroughly.

For time series analysis, it is important that the data cover the whole time range of the analysis. In addition, there should be a sufficient number of years with data, to avoid too many or long gaps. Hence, the data selection criteria at WB level have been set to:

- Data available from at least one year in the first and the last five years of the time range
- Data from at least 40 % of the years in the time range

The data selection is done separately for each of the time ranges, i.e. long time series with start year 1992 and short time series with start year 2007.

4.3 Aggregation and gap filling

4.3.1 Aggregation from site to water body

Aggregation from site to water body (WB) can be done by averaging across all monitoring sites per year, as is done in WISE Statistics (see chapter 2.2). While this is a simple approach, it can give highly misleading results for time series, especially if the number of years with data per site is not balanced and if different sites have clearly different concentration levels. The results can be particularly misleading if one site for example has been sampled in the beginning of the time period and another site (with different concentration level) in the end of the time period, with no overlap. Differences in time series length could be solved by applying data selection criteria and gap filling at monitoring site level, as explained for WB data in the previous section. However, this would exclude too much relevant data.

In the following section, the current approach for estimating annual values per WB and estimating values in years with no data (gap filling) is described. The approach involves three slightly different methods/steps, depending on the number of sites per WB and whether the data from the sites are significantly different. The method applies generalised additive models (GAM), using the free software R (R Core Team, 2020). GAMs are flexible regression models that are well suited for modelling non-linear data like time series.

A GAM model with a smooth term for year and site as a random factor will reduce the bias caused by combining site time series of different lengths. This type of model assumes that there is a common, underlying effect of year, i.e., some similarity in the time trend among sites in the same WB (Pedersen et al. 2019). This is a reasonable assumption, since sites within the same WB will be subject to at least some of the same pressures. The model estimates the common "year effect" (time trend), and a random deviation from this overall effect for each site (the random effect). The deviation is estimated as differences in the intercepts, i.e., difference in overall concentration level. The model is specified as:

$$y(year, site) = b_0 + s(year) + [site]_j + \varepsilon_i$$

where y is the observed concentration. In the notation of R's mgcv package (Wood, 2017):

```
library(mgcv)
gam(NO3 ~ s(year, k = k) + s(site, bs = 're'), data = data_waterbody, family =
Gamma(link = "log")
```

where 'site' is a factor variable and k is the "wiggliness" of the curve, i.e. how close it is allowed to follow the data. A k-value of 7 is used, which, by visual inspection of the model fits, gave a reasonable compromise between capturing the overall trend and avoiding overfitting. The part "bs = 're' ensures that a random effect is used for site. Instead of assuming a normal distribution, a gamma distribution with a log-link is used for the mean-variance relationship. The gamma distribution is well suited for modelling concentrations, which are non-negative values that often have a right-skewed distribution. A gamma distribution was also used in the gam model without a random effect of site, which is described on the next page. Values are transformed back to linear scale after prediction.

This model allows for prediction of values per year for each site, given that the random site effect is significant ($p < 0.05$). This is shown for a WB with two sites in Figure 1. Here, the yellow open dots are the predicted values for the yellow site, and the pink open dots the predicted values for the pink site. To estimate a single value per year for the WB (shown as red stars), the site predictions are averaged. If a simple average of the raw site data per year had been applied in this example, the data for the WB would have shown a sudden increase around 2010, when the yellow site was introduced. This would have been a result of adding a new monitoring site with higher concentrations, and not a reflection of the overall trend for the WB. The mixed GAM approach avoids such misleading patterns in assuming that the two sites follow the same underlying trend over time for the years without data.

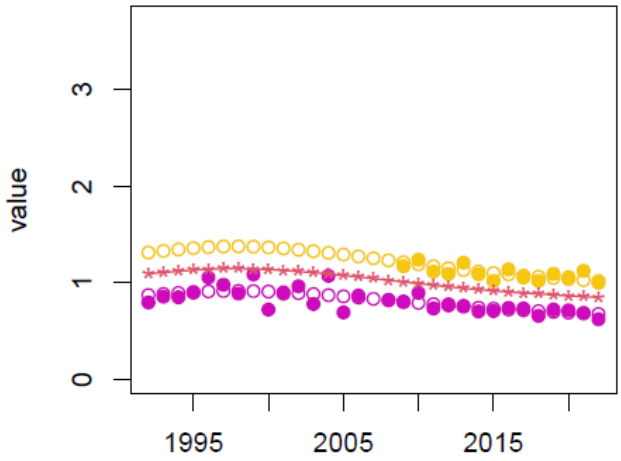

Figure 1. This WB has data from two sites (a yellow and a pink site), where the concentrations are considered significantly different by the GAM. The model then predicts one time series per site, shown as yellow and pink open dots. The overall WB time series is calculated by averaging the two site time series (red stars).

A mean value is also calculated for the years without data. This approach gives a dataset with values for each year per WB, with the GAM predictions used to estimate values also in years with missing data. This way of aggregation and filling the gaps removes some of the variation in the raw data, as it fits a smooth curve to the values over time. However, the advantages are that it reduces the impact of extreme outliers and yields a complete time series per WB representing the smoothed trend over time.

The model fits are generally good, as judged by visual inspection of the observed vs. predicted values. In rare cases, typically if data from one site are vastly different from the other site(s), the predicted values can be way outside the range of the raw data. This will again lead to misleading average values for the WB. To avoid this issue, *data from a site are excluded if any of the predicted values for the site exceed three times the maximum value among the observed data from the WB*. This affects very few sites per determinand/water category. The estimated values for the given WB will then be based on the data from the remaining sites. In the rare case that a WB no longer meets the data selection criteria after the site exclusion, it is excluded from the analysis.

In many of the WBs with more than one site, the GAM model estimates no significant site effect ($p > 0.05$ for the random site effect), indicating that the sites are similar with respect to

concentration levels and hence can be treated as a single site. In these WBs, an ordinary GAM with value as a function of year is fitted, pooling data from all sites. In R notation, that is:

```
gam(NO3 ~ s(year, k = k), data = data_waterbody, family = Gamma(link = "log"))
```

The gam model is used to predict annual values used as estimates of overall annual values for the WBs. As above, the predicted values in the years without data are used to fill the gaps in the dataset, creating a complete time series per WB. An example of this approach is shown in **Figure 2**.

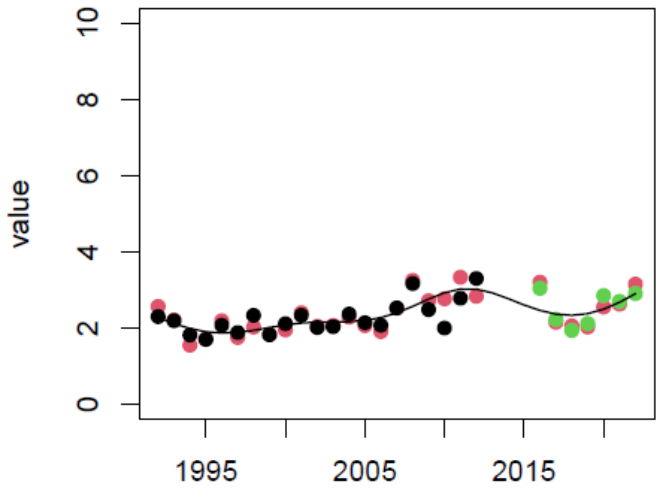

Figure 2. This WB has data from three sites (black, red and green), but the sites are not considered significantly different by the GAM model. Hence, a single GAM is fitted to the combined dataset. The predicted values are shown as a black line.

For WBs with only one site, there is no need to combine data and the annual site means could in principle have been used directly. Further, one could e.g. have used linear interpolation to estimate values in years with no data to create complete time series. However, since GAM is used to estimate annual values and complete time series for WBs with > 1 site, and also allows longer gaps in the time series, it was decided to use a similar approach here. This ensures similar structure and level of variation in the WB time series, independent of the number of sites. Hence, for WBs with only one site, the same approach as above is applied, i.e. fitting an ordinary

GAM with value as a function of year and using the model to predict values for all years (see example in **Figure 3**).

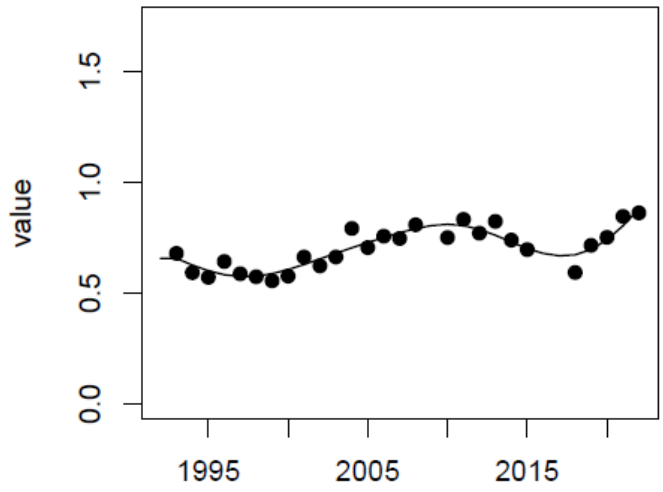

Figure 3. For WBs with only one site, a GAM is fitted to the data. The model is used to predict values for the entire time range (black line).

For all three approaches described above, data going beyond the selected time range is used when fitting the GAMs for the short time series (start year 2007), i.e. data from years before 2007 are included in the modelling if available. Since the selection criteria only require data from at least one year in the first five-year period (i.e. 2007-2011 for the short series), including data from before 2007 can help guide the model to more appropriate fits for the beginning of the time series. However, sites with data *only* from before 2007 are not included, as this can give misleading results.

If the first year with data is after 2007 for short series, or after 1992 for long series, the model fit is extrapolated back to 2007 or 1992, respectively, by setting the earlier years' values equal to the predicted value in the first year with data (see example in **Figure 4**). The same is done at the end of the time series if the last year with data is earlier than the last year of the analysis time range. If the GAM had just been extrapolated back- or forwards to the start or end year, extrapolated data would often be way out of range of the original data. This is especially the case for GAM curves changing rapidly at the ends of the time series. For WBs with significantly different site time series, it is the overall WB time series that is extrapolated.

Data reported directly as aggregated to WB level (only relevant for groundwater nitrate) are included for WBs and years where there are no other data. These data are then treated as site data in the GAM modelling.

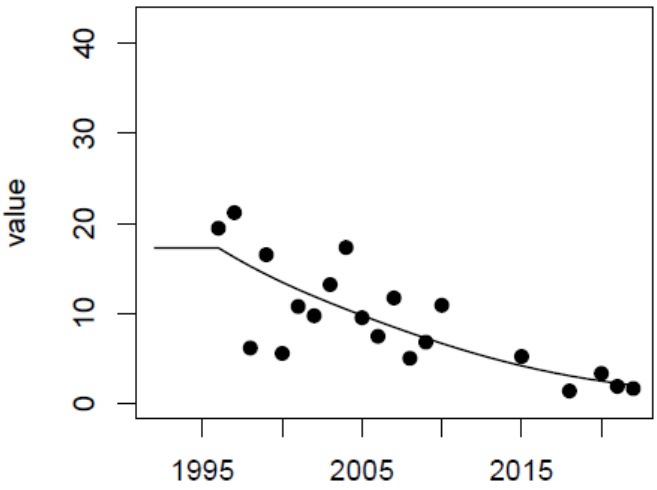

Figure 4. Example of a long time series (start year 1992) where the first observation is from 1996. Instead of extrapolating the GAM (black curve) back to 1992, estimates for 1992-1995 are set equal to the value in 1996.

All the three methods described above are coded in R and can be run automatically and fast using scripts. After the automated model fitting, all the WB time series are plotted with added GAM fits (like the examples in Figure 1 to Figure 4 above) and visually inspected for quality assurance. This is important, since the model fits in rare cases can give misleading results which may have consequences for the next steps in the analysis (calculation of country averages and overall EU trend). Bad model fits are, however, usually a result of bad input data, e.g. extreme outliers that are not detected in the initial data quality control. When this is observed, it is possible to flag such outliers and exclude them from the analysis (see chapter 2.3). *Other reasons for bad model fit can be sites with only one year of data strongly affecting the result (in case the site is excluded) or special combination of site results giving misleading results (in case the WB is excluded).* This affects only a very limited number of sites and WBs.

4.3.2 Aggregation to country level

The gap filled time series per WB are aggregated to country time series by averaging across all WBs per year. As the time series are gap filled, the number of WBs per year is the same. Separate aggregations are made for the sets of WBs with valid long (start year 1992) and short (start year 2007) time series. The country averages are further used to calculate a weighted average for Europe, which is described in the next section.

5 Time series for Europe

The summarised result for Europe is calculated as a weighted average based on the country time series and total water "size" (total length of rivers (km) or total area of lakes or groundwater (km^2) per country). This is done to avoid bias towards countries with a high density of water bodies (WBs) with valid time series. However, applying this approach there is a risk of giving too much weight to large countries with a low density of monitored WBs which are likely not to be representative for the country as a whole. Hence, there is a need to exclude countries with very low sampling density before the EU-wide calculation.

5.1 Country exclusion

To evaluate the density of WBs with valid time series, the sampling density in each country *i* is calculated as

Sampling density$_i$ = n WBs$_i$ / water size$_i$,

where n WBs is the number of WBs in country *i* with valid time series, and water size is the total river length (km) or lake/groundwater area (km^2) in country *i*, as summarized from the spatial data (see chapter 2.4).

The selection of exclusion criteria is a trade-off between including as many countries as possible while still ensuring a reasonable sampling density. The criteria should also be objective and the same across determinands within the same water category and across time ranges. An attempt was made to use the same approach as for the pesticides indicator. Here, the 5th percentile of the sampling density across all countries per year and WB category is calculated, and then the mean across all years is used as a single threshold per WB category. However, with the specific WB selection for nutrient time series and the two different time series ranges (with fewer countries and WBs for the longer range), it proved impossible to find a single percentile that could be used across WB categories and time ranges.

For nutrients it was therefore decided to use fixed thresholds. It turned out that when setting thresholds that were sufficiently relaxed not to exclude countries with a reasonable sampling density, sometimes countries with only one WB time series were included. An additional criterion on total number of WBs was therefore included.

Countries are excluded if they meet the following criteria with respect to sampling density and number of WBs:

- Rivers: Sampling density < 0.00005 WBs/km or <2 WBs
- Lakes: Sampling density < 0.0001 WBs/km^2 or <2 WBs
- Groundwater: Sampling density < 0.00004 WBs/km^2 or <2 WBs

In addition, countries/water categories for which spatial data were missing must be excluded, as this information is needed for producing European results (se next section).

5.2 Weighted average

After exclusion based on sampling density, weighted average concentrations per year ($C_{w,\,year}$) for Europe are calculated based on the country averages and total water size per country using the following formula:

$$C_{w,year} = \frac{\sum_{i=1}^{n} S_i C_i}{\sum_{i=1}^{n} S_i}$$

where S_i is the size value (total river length in km or lake/groundwater area in km^2) in country *i* and C_i is the average concentration in country *i* for the given year. Weighted averages are calculated both for the short (start year 2007) and long (start year 1992) time series.

5.3 Confidence interval

Confidence intervals for the weighted average concentrations are calculated so that they reflect the variability between the country time series. There is no consensus on how to calculate the standard error of a weighted mean, but the chosen approach seems to be frequently referred to (Gatz and Smith, 1995a, b). The third formula on page 1886 in Gatz and Smith (1995a) is applied to calculate the standard error of the mean (SEM). Subsequently, a 95% confidence interval is calculated according to the formula $C_{w,year} \pm 2 \times SEM$.

6 Trend analysis

Trends are analysed by the Mann-Kendall method in the free software R (R Core Team, 2020), using the wql package (Jassby et al., 2017). This is a non-parametric test suggested by Mann (1945) and has been extensively used for environmental time series (Hipel and McLeod, 2005). Mann-Kendall is a test for a monotonic trend in a time series y(x), which in this analysis is nutrient concentration (y) as a function of year (x). The test is based on Kendall's rank correlation, which measures the strength of monotonic association between the vectors x and y. In the case of no ties in the x and y variables, Kendall's rank correlation coefficient, tau, may be expressed as tau = S/D where S = sum_{i<j} (sign(x[j]-x[i])*sign(y[j]-y[i])) and D = n(n-1)/2. S is called the score and D, the denominator, is the maximum possible value of S. The p-value of tau is computed by an algorithm given by Best and Gipps (1974).

The tests reported are two-sided (testing for both increasing and decreasing trends). Data series with p-value <0.05 are reported as significantly increasing or decreasing, while data series with p-value >= 0.05 and <0.10 are reported as marginally increasing or decreasing. The results are summarised by calculating the percentage of WBs within each category relative to all WBs within the specific aggregation (Europe or country). The test analyses only the direction and significance of the change, not the size of the change.

The size of the change is estimated by calculating the Sen slope (or the Theil or Theil-Sen slope) (Theil, 1992; Sen, 1968) using the R software. The Sen slope is a non-parametric method where the slope m is determined as the median of all slopes (yj−yi)/(xj−xi) when joining all pairs of observations (xi, yi). Here the slope is calculated as the change per year for each WB. This is summarised by calculating the average slope (regardless of the significance of the trend) for all WBs in Europe or a country. For the relative Sen slope (Sen slope %), the slope joining each pair of observations is divided by the first of the pair before the overall median is calculated and multiplied by 100. Again, this is summarised for Europe or individual countries by averaging across WBs.

The Mann-Kendall method or the Sen slope will only reveal monotonic trends and will not identify changes in the direction of the time series over time. Hence a combination of approaches can be used to describe the time series: a visual inspection of the time series, describing whether the general impression is a monotonic trend, no apparent trend, clear shifts in direction of the trend or high variability with no clear direction; an evaluation of significant versus non-significant and decreasing versus increasing monotonic trends using the Mann-Kendall results; an evaluation of the average size of the monotonic trends using the Sen slope results.

For the trend analysis the same time series are used as for the aggregated time series plots, but without gap filling. This means that years where there are no data for any of the monitoring sites in the WB are omitted. The Mann-Kendall analysis can handle gaps in the time series. Using the gap filled values would wrongly increase the number of significant trends. However, since the trend analysis is run on GAM modelling output, the data is smoothed relative to the raw data. This also makes it more likely to get significant trends. This is a consequence of basing the trend analysis on data aggregated to water body level.

7 Present state analysis

7.1 Annual data per water body from WISE Statistics

The basis for the present state analysis is the annual mean concentrations per water body (WB) in the [AggregatedDataByWaterBody] table in WISE Statistics (WISE_Indicators in Discodata). These are averages of the annual values per monitoring site for each WB from the [AggregatedData] table. Data reported as aggregated to WB (groundwater nitrate) are included in the [AggregatedDataByWaterBody] table when reported disaggregated (samples) or aggregated data (annual data per site) are not available. For reported data aggregated to WB, handling of LOQ and corrections are the same as for reported aggregated data (see section 3.3).

Aggregation to WB level based on the annual values per monitoring site prevents bias towards sites with more samples per year in the summary calculations. This is also in line with the Data Dictionary. The aggregation to WB follows the same principles as the aggregation from samples to annual values per monitoring site:

- The annual mean values per WB are calculated from annual mean values per site within the WB. The annual site mean values are used directly whether they are set as being below LOQ or not. This is because half LOQ has already been used for the disaggregated samples that were below LOQ in the calculation of the site mean values.
- If any of the annual site mean values are set as below LOQ, the LOQ per WB is set to the highest of the LOQs set for these means. Otherwise, the aggregated LOQ is set to the highest across all the annual site LOQs.
- Whether the mean is below LOQ is evaluated against the LOQ per WB. If the mean is below the LOQ per WB it is set equal to the LOQ per WB. There is an exception where all sample values are above the aggregated LOQ, while the mean is below the LOQ per WB. Then the original mean value is kept and the mean is set as not being below LOQ.
- The number of samples and number of samples below LOQ are calculated as the sum of the information for the individual sites, so these represent the total number of individual samples.

For duplicates the same principles apply as for annual values per monitoring site (see section 3.5). Data aggregated from site data are given preference to data reported at WB level.

7.2 Calculations for the present state analysis

In the present state analysis, average WB concentrations are calculated across the last six years with data. With such a short time period and for assigning one concentration level per WB, any variability in number of sites per WB between years is considered not important. Hence the regular WB averages in the [AggregatedDataByWaterBody] table can be used rather than aggregated values from GAM modelling. This reduces complexity and avoids introducing

estimated values. The 6-year average is used to remove some inter-annual variability. It also gives more WBs, since data are not available for all WBs each year. This means that any WB with at least one data point within the last six years can be used in this analysis, which gives far more WBs than in the time series analysis. Until 2024 three years were used but extending it to six years makes it possible to include WBs with a monitoring cycle of six years (e.g. from WFD surveillance monitoring).

For non-EU countries where WBs have not been defined, monitoring site data from the [AggregatedData] table are used as basis for the 6-year average. Individual monitoring sites without WB information from countries where WBs have been defined (from within and outside EU) are not included in the analysis.

The WBs (and sites, where relevant) are split into quintiles based on the distribution of the 6-year average concentrations, and the results are summarised per country as percentage of WBs per quintile class. As the total set of WB averages changes between years, the quintile thresholds will also change marginally. However, the purpose of the analysis is to compare the distribution of concentrations among countries. The results from consecutive analyses are never compared.

Note that nutrient concentrations vary naturally. E.g. slow-flowing lowland rivers will naturally have higher nutrient concentrations than alpine rivers. Consequently, a lower proportion of water bodies in the lowest concentration classes does not necessarily imply higher anthropogenic pollution pressure. The natural variability is the reason why quintile classes are used rather than fixed thresholds that may indicate status of the WBs. Different water body types have different thresholds, and using common thresholds across types would lead to wrong conclusions.

List of abbreviations

Abbreviation	Name	Reference
EEA	European Environment Agency	www.eea.europa.eu
RBMP	River Basin Management Plan	
WB	Water body	
WFD	Water Framework Directive	https://environment.ec.europa.eu/topics/water/water-framework-directive_en

References

Best, D., and Gipps, P., 1974, *Algorithm AS 71: The Upper Tail Probabilities of Kendall's Tau*, Journal of the Royal Statistical Society, Series C (Applied Statistics), 23(1), 98-100.

Gatz, D. F., Smith, L., 1995a, *The standard error of a weighted mean concentration–I. Bootstrapping vs other methods,* Atmospheric Environment 11, 1185-1193.

Gatz, D. F., Smith, L., 1995b, *The standard error of a weighted mean concentration–II. Estimating confidence intervals,* Atmospheric Environment 29, 1195-1200.

Hipel, K., McLeod, A., 2005, *Time Series Modelling of Water Resources and Environmental Systems, Volume 45 - 1st Edition*, (https://www.elsevier.com/books/time-series-modelling-of-water-resources-and-environmentalsystems/hipel/978-0-444-89270-6) accessed March 25, 2022.

Jassby, A. D., Cloern, J. E., Stachelek, J., 2017, *Exploring water quality monitoring data*, (https://cran.r-project.org/web/packages/wql/) accessed March 25, 2022.

Mann, H. B., 1945, *Nonparametric Tests Against Trend*, Econometrica 13(3), 245–259 (https://www.jstor.org/stable/1907187) accessed March 25, 2022.

Pedersen, E. J., Miller, D. L., Simpson, G. L., Ross, N., 2019. *Hierarchical generalized additive models in ecology: an introduction with mgcv*. PeerJ 7:e6876 https://doi.org/10.7717/peerj.6876, accessed December 12, 2024.

Theil, H., 1992, *A Rank-Invariant Method of Linear and Polynomial Regression Analysis*, in: Raj, B., Koerts, J. (eds), Henri Theil's Contributions to Economics and Econometrics: Econometric Theory and Methodology, Springer Netherlands, Dordrecht, pp. 345–381.

R Core Team, 2020, *R: The R Project for Statistical Computing*, (https://www.r-project.org/index.html) accessed March 25, 2022.

Sen, P. K., 1968, *Estimates of the Regression Coefficient Based on Kendall's Tau*, Journal of the American Statistical Association 63(324), 1379–1389 (https://www.tandfonline.com/doi/abs/10.1080/01621459.1968.10480934) accessed March 25, 2022.

Wood, S. N., 2017, *Generalized Additive Models: An Introduction with R (2nd edition),* Chapman and Hall/CRC.