

JAMP Guidance on input trend assessment and the normalisation of loads

Revision made by NIBIO (Norway) and SLU (Sweden) 2017

(Agreement number: 2003-09)[[1]](#footnote-1)

# Contents

[Contents 1](#_Toc486936653)

[1. Objective 2](#_Toc486936654)

[2. General considerations 2](#_Toc486936655)

[2.1 Initial analysis 2](#_Toc486936656)

[2.2 Consideration of overall reliability 3](#_Toc486936657)

[2.3 General procedure 3](#_Toc486936658)

[3. Normalisation of loads and assessment of trends 4](#_Toc486936659)

[3.1 General introduction on normalisation 4](#_Toc486936660)

[3.2 Normalisation and trend assessment for annual riverine data 5](#_Toc486936661)

[3.2.1. Normalisation and trend assessment if trends can be assumed to be linear 5](#_Toc486936662)

[3.2.2. Normalisation and trend assessment if trends are nonlinear 5](#_Toc486936663)

[3.2.3. Normalisation for short series 5](#_Toc486936664)

[3.2.4. Mann-Kendall tests for annual data 6](#_Toc486936665)

[3.2.5 Example 6](#_Toc486936666)

[3.3 Normalisation and trend assessment for monthly or more frequent riverine data 9](#_Toc486936667)

[3.3.1. Normalisation and trend assessment if trend can be assumed linear 9](#_Toc486936668)

[3.3.2. Normalisation and trend assessment if trends are nonlinear 9](#_Toc486936669)

[3.3.3 Mann-Kendall tests for seasonal data 9](#_Toc486936670)

[3.4 Normalisation of total inputs 10](#_Toc486936671)

[3.5 Quality control for normalisation and trend models 10](#_Toc486936672)

[4. Recommendations for normalisation and trend analysis 11](#_Toc486936673)

[5. Software 11](#_Toc486936674)

[6. References 11](#_Toc486936675)

# 1. Objective

1. The objective of the Comprehensive Study on Riverine Inputs and Direct Discharges (RID (see revised RID-principles; Agreement 2014-04) is to assess, as accurately as possible, all riverborne and direct inputs of selected pollutants to Convention waters on an annual basis. Inputs are a function of concentration measurements of a substance and of flow rate. Inputs are influenced to a greater or lesser extent by factors such as flow and suspended particulate matter (SPM). Under the OSPAR Convention, the purpose of trend assessment of inputs is to identify and quantify any significant patterns of change in inputs in order to support the relevant OSPAR Strategies. While there is a need to consider both actual and normalised inputs in any overall assessments, normalised inputs are more likely to be important in determining the effectiveness of measures at source within the OSPAR catchment or river basin area.

2. The aim of this document is to provide overall guidance on a procedure to identify and quantify temporal trends in inputs, including normalisation.

# 2. General considerations

## 2.1 Initial analysis

3. Before starting the proposed trend assessment procedure, the following factors need to be taken into account:

**a. data quality**

Generally, it should be guaranteed that all data have been checked for analytical or transfer errors.

**b. length of time series**

For the application of statistical methods, the time series should as a rule of thumb comprise measurements/estimates over a period of at least ten years. Failing to do so greatly enhances the probability of producing a “trend” that is only due to climatological or other random variation.

**c. measurements less than the LOD or LOQ**

All the concentration measurements falling below the limit of detection (LOD) or limit of quantification (LOQ) should be replaced with appropriate substitute values in preparation for normalisation and trend analysis; preferably <20% of observations of the relevant limit of detection or quantification for this purpose. However, a prerequisite of proper load estimation is that the impact of measurements below LOD and/or LOQ is not too large.

**d. changes over time in LOD or LOQ**

A high percentage of measurements below LOD and/or LOQ seriously affects the modelling of the load and/or concentration – run-off functions and hence the normalisation step. The modelling results become even less reliable if there are several changes of the LOD/LOQ over the time series. In order to avoid spurious trends when there are measurements below LOD/LOQ:

(i) LOD/LOQ should be constant over the whole time series for automated trend assessments.

(ii) Generally, trend assessment should not be done when values below LOD/LOQ in the same time series are based on different levels of detection. Alternatively, all LOD/LOQ should be normalised to the highest LOD/LOQ in the series before proceeding.

**e. missing values and outliers**

A graphical presentation provides a visual summary of the general trend in the data, to reveal extreme values, and an opportunity to reveal missing values and other inconsistencies in the data. Outliers can be detected using an appropriate test, as recommended by ICES (2000). Missing values will narrow the scope of the methods available for normalisation. Procedures for dealing with missing values should be justified for each application.

**f. exceptional events, e.g. one-off accidents**

 Single values may seriously affect the quality of the flow normalisation. It is, therefore, necessary to inspect the raw data and to flag questionable values. If it is clear that the values are the result of an exceptional event, they can be omitted from the trend analysis. However, any such omission should be justified explicitly in the report.

## 2.2 Consideration of overall reliability

4. Both the reliability of the basic data and the statistical-assessment techniques used to evaluate any trend have an impact on the reliability of the overall assessment. It is, therefore, generally important that the assessor has a broad appreciation of the circumstances, including any incidents or changes that could distort the long-term pattern of change, relating to the data. It is also important that the assessor should apply the techniques of assessment with understanding.

5. Where data have been reported to OSPAR under the RID programme, they should have been subject to the standards of quality assurance and verification detailed in the RID Principles. Where data are used which have not been subject to OSPAR reporting procedures (e.g. data sets before 1990), the assessor needs to ensure that the data are sufficiently reliable for use. This implies that basic standards (such as those established by the ISO) have been complied with for sampling, analysis and data handling.

6. Where the data are reliable and input estimates are consistent it is likely that the trend assessment will be reliable. Where there is a need to deal with gaps in information, to normalise for variations or anomalies or for varying LOD/LOQ, the assessment may be less reliable or have lower power. It is, therefore, important that any steps taken to deal with gaps, etc. are duly justified in the assessment report and unnecessary changes in methodology or LOD/LOQ are avoided in the collection of data. An estimation of the effect of any normalisation made on the reliability of the trend assessment should be quantified. In respect of this, some form of sensitivity analysis is advised.

## 2.3 General procedure

7. Generally, normalisation for fluctuations in run-off should take place using the raw data, following the sequence given in Figure 1. The normalised loads should then be compiled into annual loads before undertaking trend analysis on the basis of the normalised annual loads.

8. If only annual data sets are available, the simplified approach explained in chapter 3.2 should be used.

| **1. Provision of raw data** |
| --- |
|   |
|  |
| **2. Normalisation and trend assessment of seasonal loads** |
|  |
|  |
| **3. Aggregation of the normalised loads** |
|  |
|  |
| **4. Trend assessment of annual loads** |
|  |

*Figure 1: Conceptual approach for normalisation and trend analysis*

9. Temporal trends can be assessed on seasonal data (e.g., monthly) or on the basis of annual (aggregated) values. An automatic process should be avoided considering that there often are outliers that have strong influence both on the normalisation step and the trend analysis and that some series cannot be normalised due to weak relationships with flow.

10. Moreover, autocorrelations can be an issue, especially for seasonal data, and p-values can be underestimated if autocorrelations are not taken into account. If substantial autocorrelation is expected the following solutions are feasible:

- only use trend analysis for aggregated annual data or

- incorporate autocorrelation estimates in the model

# 3. Normalisation of loads and assessment of trends

## 3.1 General introduction on normalisation

11. It has long been recognised that normalisation of the collected data can greatly facilitate comparisons over time. Different types of regression techniques have been employed to remove or suppress natural fluctuations in time series of air and water quality data (Thompson *et al*., 2001; Uhlig & Kuhbier, 2001; Stålnacke & Grimvall, 2001).

12. In this guideline, we address the fact that riverine loads of substances can vary strongly with run-off, and that natural fluctuations in run-off can conceal or distort important trends in the human impact on nutrient inputs to the sea. In principle, the recommended procedures can also be applied to explanatory variables other than flow. For example, it might be desirable to normalise the load of substances bound to particles with respect to the load of suspended particulate matter.

13. Flow-normalisation requires knowledge about the statistical relationship between flow and concentration, as well as between flow and load.

14. Loads often are computed by multiplication of concentrations with flow, the relationship between load and flow can in many circumstances be assumed to be linear. Therefore statistical normalisation models are often based on regression models.

15. If the available series are short the fitting of a model might not be feasible. In that case a simple division between load and flow can still be better than not conducting any normalisation at all.

## 3.2 Normalisation and trend assessment for annual riverine data

### 3.2.1. Normalisation and trend assessment if trends can be assumed to be linear

16. A basic normalisation and trend assessment model for loads can be written as

$$L\_{i}=α+β\_{1}t\_{i}+β\_{2}q\_{i}$$

with $L\_{i}$ the load for year i, $q\_{i}$ the flow for year i and $β\_{1}t\_{i}$ a linear temporal trend.

The estimate for $β\_{2}$ can be used to determine if the influence of flow on loads is statistically significant or not, i.e. if normalisation is necessary or not.

The parameter $β\_{1}$ can be tested against the null hypothesis of no linear change in time. If $β\_{1}$ is significantly different from 0 there is a significant linear trend in the data. The estimate of $β\_{1}$ also quantifies the mean increase of decrease per year.

17. From this model normalised loads can be computed as

$$NL\_{i}=L\_{i}-β\_{2}q\_{i}+β\_{2}\overbar{q}$$

where $\overbar{q}$ is a long term average for flow.

### 3.2.2. Normalisation and trend assessment if trends are nonlinear

18. In longer time series it is often difficult to motivate the presence of a linear temporal trend in data. Instead both increases and decreases can be observed during different time periods or a trend is observed during few years only with rather constant levels before and/or after. In that case the normalisation model should be normalised by incorporating a smooth trend component:

$$L\_{i}=α+f\left(t\_{i}\right)+β\_{2}q\_{i}$$

19. The trend component can be fitted by any smoother, e.g. loess or spline smoothing (Hastie and Tibshirani, 1990, Stålnacke and Grimvall, 2001).

20. The normalisation model is still based on a linear relationship between flow and load and again $β\_{2}$ can be used to determine if the influence of flow on loads is statistically significant or not.

21. Normalised loads can be computed in the same way as for the normalisation model with a linear trend.

22. The smooth trend model does not contain any parameter that can be used to test for trend. Instead it is recommended to conduct a Mann-Kendall test (section 3.2.4) on the normalised values. The magnitude of the trend can also be quantified as percentage change compared to initial levels in the series.

### 3.2.3. Normalisation for short series

23. If regression-based normalisation is not feasible due to short series a ratio-normalisation can be conducted:

$$RL\_{i}=\frac{L\_{i}}{q\_{i}}$$

24. Normalised loads can then be computed by

$$NRL\_{i}=\frac{L\_{i}}{q\_{i}}∙\overbar{q}$$

where $\overbar{q}$ is again a long term average for flow.

### 3.2.4. Mann-Kendall tests for annual data

25. If original or normalised annual data are tested for trend Mann-Kendall tests can be used.

Mann-Kendall tests are non-parametric trend tests that do not demand normal distribution of data. They are instead computed on ranks of data:

$$MK=\sum\_{j<i}^{}sgn\left(y\_{i}-y\_{j}\right)$$

where sgn(x) is the sign function

$sgn\left(x\right)=\left\{\begin{matrix}1, if x>0\\0, if x=0\\-1, if x<0\end{matrix}\right.$.

Results are presented in terms of a p-value, i.e. indicating if there is a significant trend or not. They do not provide an estimate of the magnitude of the trend, but can be combined with Theil-Sen’s slope that gives the median increase or decrease per year.

26. Mann-Kendall test can also incorporate a normalisation step and are then called **partial Mann-Kendall tests** (Libiseller and Grimvall, 2001). They are a simpler alternative if regression-based normalisation model is not applicable.

27. Mann-Kendall tests for annual data are recommended if there are at least 10 years of data.

### 3.2.5 Example

28. This procedure may be illustrated by an example. The following table contains the UK total-N nitrogen inputs data (sea area 196) for 1990-2014; see also black dots in figure 2.

| **year** | **flow** | **mean** |
| --- | --- | --- |
| 1990 | 4390 | 17.5 |
| 1991 | 2658 | 11.5 |
| 1992 | 4097 | 14 |
| 1993 | 6464 | 19.3 |
| 1994 | 5960 | 21.5 |
| 1995 | 8195 | 23.7 |
| 1996 | 3339 | 13.5 |
| 1997 | 2013 | 7.3 |
| 1998 | 6995 | 22.1 |
| 1999 | 7730 | 21.0 |
| 2000 | 10425 | 27.0 |
| 2001 | 13510 | 30.6 |
| 2002 | 10108 | 23.9 |
| 2003 | 7464 | 19.8 |
| 2004 | 5482 | 14.7 |
| 2005 | 2549 | 9.2 |
| 2006 | 4615 | 11.5 |
| 2007 | 6887 | 12.4 |
| 2008 | 11795 | 28.0 |
| 2009 | 5595 | 15.3 |
| 2010 | 5414 | 13.6 |
| 2011 | 3010 | 9.6 |
| 2012 | 8160 | 20.8 |
| 2013 | 6838 | 19.6 |
| 2014 | 13525 | 29.8 |

A linear trend model is fitted:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Value** | **Standard error** | **t value (Value/St. Err)** |  **p-value** |
| Intercept (α) | 418 | 109.6 | 3.81 | 0.00095 |
| Year (1) | -0.207 | 0.055 | -3.76 | 0.0011 |
| Run-off (β2) | 0.00205 | 0.000125 | 16.5 | 0.00000 |

29. This corresponds to the regression equation

Load = 418 – 0.207 ∙ year + 0.00205 ∙ run-off

30. The impact of the run-off on the load can be described as β2 = 0.00205. This is highly significant because the corresponding p-value is below the 5% significant level. The conclusion in this example, therefore, is that normalisation should be carried out.

31. The trend in total nitrogen can be described as linear and is then tested in the same model. The parameter 1 = -0.207 and indicates a linear negative trend that is significant with a p-value of 0.0011.

32. A smooth trend is an alternative as shown in Figure 2 below. It can be argued that the smooth line represents data in the end of the series better. If a smooth trend is used in the normalisation model, we get the following results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Value** | **Standard error** | **t value (Value/St. Err)** | **P value** |
| Intercept (α) | 4.508 | 0.879 | 5.126 | 0.000046 |
| Run-off (β2) | 0.00206 | 0.000121 | 17.08 | 0.00000 |

33. The normalisation model is very similar to the one with a linear trend. The intercept has, however, another interpretation than the traditional regression model.

34. To determine the significance of the non-linear trend a Mann-Kendall test can be computed on the normalised values.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Value** | **Variance** | **P value** |
| Mann-Kendall test | -140 | 1833.33 | 0.0012 |

35. The Mann-Kendall test indicates a significant negative trend with a p-value of 0.0012.

 

*Figure 2: observed values in black and normalised values with linear (in red) and smooth trend (in blue).*

36. Instead of doing normalisation in a regression model we could also use the partial Mann-Kendall test.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Value** | **Variance** | **P value** |
| Partial Mann-Kendall test | -33.54 | 280.4 | 0.052 |

37. In this case, with long series and a clear relation between flow and load, the partial Mann-Kendall test works not as well as regression-based normalisation. Still the result is close to significant, whereas the result of a Mann-Kendall test on the original loads is far from significant, indicating that the normalisation for flow works reasonably well.

## 3.3 Normalisation and trend assessment for monthly or more frequent riverine data

38. For monthly or more frequent data, similar procedures can be applied. It is, however, necessary to include seasonal variations in the normalisation model.

### 3.3.1. Normalisation and trend assessment if trend can be assumed linear

39. Models for seasonal data have the same structure as models for annual data, but contain also an estimation of the seasonal components:

$$L\_{ij}=α+β\_{1}t\_{i}+seasonal component\_{j}+β\_{2}q\_{ij}$$

Seasonality can be modelled in different ways, e.g. with seasonal mean values, a smooth cyclic function over the year, or a sine-cosine function.

40. Significance testing for the flow-load relationship and the trend can be conducted in the same way as for annual data.

41. It should also be considered if either the flow-load relationship or the trends vary over the seasons. In that case the model can be normalised to allow for different relationships for different seasons or months, i.e. by estimating different parameters $β\_{2j}$. Also different linear trend lines can be estimated using different parameters 1j.

$$L\_{ij}=α+β\_{1j}t\_{ij}+seasonal component\_{j}+β\_{2j}q\_{ij}$$

42. However, it needs to be considered that the model can easily be over-parametrised. Seasonal trends and relationships should only be incorporated if there is a strong expectation that they are needed.

### 3.3.2. Normalisation and trend assessment if trends are nonlinear

43. Equivalently, the smooth trend model can be normalised to seasonal data, leading to the model:

$$L\_{ij}=α+f\left(t\_{i}\right)+seasonal component\_{j}+β\_{2}q\_{ij}$$

44. Trend testing can again be done by Mann-Kendall tests and the magnitude of change can be quantified as percentage change compared to initial levels in the series.

### 3.3.3 Mann-Kendall tests for seasonal data

45. Seasonal Mann-Kendall tests (Hirsch and Slack, 1984) can be used for seasonal data, by computing individual Mann-Kendall tests for each season. Autocorrelations are then also taken into account. The seasonal trend tests can be interpreted separately or be combined to a single test representing the entire series.

46. Also partial Mann-Kendall test can be used on seasonal data.

## 3.4 Normalisation of total inputs

47. Normalisation procedures are not applicable for direct inputs. In order to calculate normalised total inputs, the following procedure is recommended:

Examine whether normalisation of the riverine load is necessary.

* If that is the case methods from section 3.1 or 3.2 are used and direct discharges are added to riverine inputs after normalisation.
* If no normalisation is necessary, raw data for direct and riverine input are added to yield the total input.

If total inputs contain multiple riverine inputs, the normalisation should be conducted for each river separately and the normalised loads should be added to define the total normalised input

48. Total inputs should not be flow-normalised since the flow-dependence is different for riverine and direct inputs and a combined normalisation on the sum of different inputs would be misleading.

49. Direct discharges should also be assessed for temporal trends separately.

## 3.5 Quality control for normalisation and trend models

50. Visual inspection of original data and of residuals should be used to ascertain that assumptions about the form of the trend and, if applicable, the seasonal structure are fulfilled. This means that residuals can be plotted against time to ensure that no temporal structures are missed and residuals can be plotted against season to see if the seasonal component is fitted correctly.

51. Residuals should also be used to ascertain approximate normal distribution. If data distribution is very skewed, models could be fitted on log-transformed data producing log-transformed normalised values and trends on the log-scale.

52. Other properties of the series must also be considered here:

a. The relationship between flow and load should be reasonably strong. If relationships are weak it can be concluded that there are other important influencing factors which have not been taken into account and might disturb the normalisation step.

b. If there are reasons to believe that a series is not consistent in time, e.g. due to substantial changes in methodology or due to eliminated or added point sources it should be considered to analyse the different parts of the series separately in order to not introduce an artificial trend. This is in more detail discussed in Larsen and Svendsen (2013).

c. Not only loads, but also concentrations can be flow dependent due to dilution or washout effect. If such effects are suspected it could be meaningful to further analyse data in order to quantify this dependence and, if substantial, loads could be computed from normalised concentrations.

d. If time series are long it is also possible that several of the relationships in the model change over time. For example, the load-flow relationship can change over time if important point sources are eliminated. Also seasonal variation can change. Incorporating time-varying relationships in single time series is prone to lead to over-parameterisation of the model and should be considered in a multiple series context only.

# 4. Recommendations for normalisation and trend analysis

* Trend analysis can be made for individual rivers, direct discharges and total inputs.
* For total inputs and other aggregated data, the individual source emission series (each river, direct discharges and unmonitored areas) should prior to any trend analysis be carefully scrutinized to ensure completeness and identify data quality problems.
* Flow-normalisation of data should be done for individual rivers, if possible, but not on direct discharges. The exact method used for normalisation is of smaller importance.
* Observed or normalised loads of individual rivers and direct discharges can be summed to total inputs
* The trend analysis should be done on normalised data, if possible, and should always be accompanied by a graphical representation of the data.
* Significance testing of the trend should be done by an appropriate trend test. If in doubt, Mann-Kendall tests can be used in most situations.
* A quantification of the magnitude of the trend should be made, if possible.
* If there are substantial changes were made in methodology during the observation period it should be considered to analyse the series in two parts.
* Presenting time series grouped by region, by substances, or by originating country can enhance the analysis both by the possibility to identify common trends, but also to be able to identify data quality problems.

# 5. Software

There is a variety of software available to do trend analysis and normalisation. If normalisation and trend analysis are done by regression methods most statistical software can be used. The free software R (R Core Team, 2015) provides a number of different packages to compute Mann-Kendall tests as well as linear and smooth regression models. There are also specialised software to conduct normalisation with smooth trends (e.g. MULTITREND (Grimvall et al., 2009) or RTrend (Uhlig et al. 2002)) or to compute different versions of Mann-Kendall tests (e.g. MULTITEST (Grimvall et al., 2011)).

# 6. References

Hastie, T.J. & R.J. Tibshirani, 1990. Generalized Additive Models. Chapman & Hall, 335 p.

ICES, 2000. Report of the Working Group on Statistical Aspects of Environmental Monitoring (WGSAEM). ICES CM 2000/D:1, Agenda item 4.

Grimvall A., Libiseller C., and Wahlin K., 2009. MULTITREND - a Visual Basic program for joint evaluation of smooth trends. Linköpings universitet.

 [http://www.ida.liu.se/divisions/stima/research/Software/index.en.shtml](http://www.ida.liu.se/divisions/stima/research/Software/index.en.shtm)

Grimvall A., Libiseller C., and Wahlin K., 2011. MULTITEST - a VBA macro for Mann-Kendall tests of temporal trends. Linköpings universitet.

 <http://www.ida.liu.se/divisions/stima/research/Software/index.en.shtml>

Larsen, S., Svendsen, L., 2013. Statistical aspects in relation to Baltic Sea Pollution Load Compilation. Task 1 under HELCOM PLC-6 (No. 33). Aarhus University, DCE – Danish Centre for Environment and Energy.

Hirsch R.M & Slack j.R., 1984. A Nonparametric Trend Test for Seasonal Data With Serial Dependence. Water resources research 20, 6, 727-732.

Libiseller C. & Grimvall A., 2001. Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. Environmetrics 13.71-84.

Stålnacke, P. & A. Grimvall, 2001. Semiparametric approaches to flow-normalisation and source apportionment of substance transport in rivers. *Environmetrics* 12:233-250.

Thompson, M.L., Reynolds, J., Cok, L.H., Guttorp, P., Sampson, P.D., 2001. A review of statistical methods for the meteorological adjustment of ozone. *Atmos. Environ*., in press.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Uhlig, S. & P. Kuhbier, 2001. Trend Methods for the Assessment of the Effectiveness of Reduction Measures in the water system. UFOPLAN: 298 22 244 (in abbreviated version annex of INPUT 04/4/1).

Uhlig, S. et al., 2002. Development of a software package for trend assessment of loads within international working groups (OSPAR, HELCOM). Part 1: International pilot assessment. FKZ (UFOPLAN) 201 22 216/01 (also in INPUT 02/4/1-Add.1).

1. Source: ASMO 2003 Summary Record – ASMO 03/13/1, § 3.34-3.36, as amended by INPUT 2005 Summary Record - INPUT 05/8/1, Annex 8. Revision 2017 [↑](#footnote-ref-1)