STATISTICAL ASPECTS IN RELATION TO BALTIC SEA POLLUTION LOAD COMPILATION

Task 1 under HELCOM PLC-6

Technical Report from DCE – Danish Centre for Environment and Energy No. 33 2013

AARHUS UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

STATISTICAL ASPECTS IN RELATION TO BALTIC SEA POLLUTION LOAD COMPILATION

Task 1 under HELCOM PLC-6

Technical Report from DCE – Danish Centre for Environment and Energy No. 33 2013

Søren E. Larsen[1]
Lars M. Svendsen[2]

[1] Aarhus University, Department of Bioscience
[2] Aarhus University, DCE - Danish Centre for Environment and Energy

AARHUS
UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Series title and no.: Technical Report from DCE – Danish Centre for Environment and Energy No. 33

Title: Statistical aspects in relation to Baltic Sea Pollution Load Compilation
Subtitle: Task 1 under HELCOM PLC-6

Authors: Søren E. Larsen[1] & Lars M. Svendsen[2]
Institution: Aarhus University, [1]Department of Bioscience, [2]DCE - Danish Centre for Environment and Energy

Publisher: Aarhus University, DCE – Danish Centre for Environment and Energy ©
URL: http://dce.au.dk/en

Year of publication: December 2013
Editing completed: December 2013
Referee: Susanne Boutrup

Financial support: Helsinki Commission (HELCOM) Baltic Marine Environment Protection Commission has kindly provided funding

Please cite as: Larsen, S.E. & Svendsen, L.M. 2013. Statistical aspects in relation to Baltic Sea Pollution Load Compilation. Task 1 under HELCOM PLC-6. Aarhus University, DCE – Danish Centre for Environment and Energy, 34 pp. Technical Report from DCE – Danish Centre for Environment and Energy No. 33. http://dce2.au.dk/pub/TR33.pdf

Reproduction permitted provided the source is explicitly acknowledged

Abstract: HELCOM periodic pollution load compilation (PLC) assessments reports status and development in total annual runoff and total annual waterborne and airborne nutrient inputs to the Baltic Sea. This report deals with statistical methods for evaluating time series of annual runoff and nutrient inputs. Methods included are hydrological normalization of nutrient time series, trend analysis and a method for testing fulfilment of HELCOM Baltic Sea Action Plan (BSAP) nutrient reduction targets. Further is described how to fill in data gaps and to estimate the total uncertainty in nutrient inputs. These statistical methods are also included in the revised PLC guidelines.

Keywords: Statistical methods, nutrients, runoff, hydrological normalization, trend estimations, uncertainty of nutrient inputs, data gaps, outliers, trend analysis, estimation of changes, evaluation off fulfilling BSAP reduction targets, HELCOM

Layout: Graphic Group, AU Silkeborg
Front page photo: Søren Erik Larsen

ISBN: 978-87-7156-049-7
ISSN (electronic): 2245-019X

Number of pages: 34

Internet version: The report is available in electronic format (pdf) at http://dce2.au.dk/pub/TR33.pdf

Contents

1 Introduction

One of the key pressures related to the eutrophication and quality of the water of the Baltic Sea is waterborne (and airborne) nutrient inputs. In the Baltic Sea Action Plan from 2007 (BSAP 2007), eutrophication targets were set, and based on these preliminary maximum allowable inputs, country-allocated nutrient reduction targets were developed and adopted. IN HELCOM Copenhagen Ministerial Declaration from 3.October 2013 Contracting Parties decided on revised nitrogen and phosphorus input reduction targets.

The Contracting Parties of HELCOM have implemented several measures to reduce nutrient losses and discharges from both point sources and diffuse sources. Among other objectives, the periodic pollution load compilation (PLC) reports total annual runoff and total annual waterborne (and airborne) nutrient inputs to the Baltic Sea, evaluates time series of these annual values for trends, evaluates the importance of losses from different sources and the effect of different measures (HELCOM, 2013). The responsible working group for PLC assessments is the HELCOM LOAD group, and among other task:

- Evaluates the annual inputs and country-wise reductions in nutrient inputs and follows up on whether reduction targets are significantly fulfilled.
- Evaluates the data quality and ensures necessary corrections and adjustments of questionable or incomplete data.
- Considers and improves the trend analysis approach and the flow normalization procedures.
- Finalizes the criteria for assessing "reaching the BSAP targets of a country".
- Assesses the information provided by the Contracting Parties to determine whether they contribute to significantly reaching their nutrient reduction targets as defined in the BSAP.

All these tasks and objectives call for a standardized and appropriate methodology, including statistical methods related to trend analysis, to identify the extent of trends, estimate uncertainty in datasets and evaluate whether reduction targets are met to allow the most qualified decisions to be made regarding possible trends and acceptance of reduced inputs.

The HELCOM Heads of Delegation (HELCOM HOD 37/2012) adopted the project "Sixth Baltic Sea Pollution Load Compilation (PLC-6)".One of the tasks under this project is the "Development of a standardized methodology to calculate uncertainties in national datasets, including a methodology for filling in data gaps and missing data". The task included development of methods for testing if the annual inputs from the Contracting Parties are significantly reduced and for testing whether reduction targets are fulfilled.

This report describes and includes a theoretical treatment of the statistical methods to be applied in pollution load compilation assessments. Focus points are waterborne input – for the development of core input pressure indicators and for determining in the annual assessments whether the Contracting Parties fulfil reduction targets. The described methods include flow normalization of nutrient inputs, testing for trends, filling in data gaps, estimation of dataset uncertainty and, finally, how to test whether reduction targets are fulfilled. These statistical methods will be included in the revised PLC guidelines by 2014.

The statistical procedure for analysing for, especially downward, trends, in the normalized nutrient input values plays an important role in the pollution input compilations. The preparation of the data for trend analysis should include an assessment of the data quality, and in this report proposals are presented on how to fill in gaps/missing data in input time series and how to test for outliers in the data (chapter 2).

Furthermore, a study of the variability in the data sets behind the time series is important for assessing the size of the different components of variance. If some components can be reduced, the trend analysis will be more precise, and chapter 3 includes and discusses methods to estimate variance components and total uncertainty.

A final step in the preparation of the data is hydrological normalization of the yearly inputs in order to remove some of the effects of climate in the trends and to smooth out the input time series. This is described in chapter 4.

A number of different trend analysis methods, both non-parametric and parametric, exist. In former pollution input compilations, the non-parametric method based on Kendall's tau has been used. This method is known as the Mann-Kendall's trend test. Trend methods are described in chapter 5.

In chapter 6, a method for testing the fulfilment of reduction targets is presented. The method is based on a statistical test of mean values. The chapter also includes a definition of a traffic light system for inputs to determine which marine Baltic Sea sub-basins or which Contracting Parties (or catchments) fulfil, almost fulfil or do not fulfil the reduction requirements (or input ceilings), as outlined in document 5/2.2 from HELCOM LOAD 4/2012.

In chapter 7, we illustrate the proposed methods by a step-by-step analysis of real input data from the PLC water database to exemplify the practical use of the proposed methodologies.

In a concluding chapter, chapter 8, we discuss the different methods presented for normalizing, trend testing and estimating variance components, filling gaps, and testing the fulfilment of reduction targets. We provide recommendations on which methods to use for the different statistical tasks involved in preparing pollution load compilations.

The report concludes with an annex including an in-depth mathematical treatment of the Mann-Kendall trend test. Mathematical symbols are defined and described in the relevant sections of the report.

The authors want to express their gratitude for the funding provided by Helsinki Commission (HELCOM Project No. 11.43) approved by HELCOM HOD 37/2012, cf. LD 66. Further we want to thanks participants of the PLC6 project and HELCOM LOAD for comments and inputs, and HELCOM Secretariat for help and support.

2 Data gaps and outliers

The reliability of all statistical methods and statistical analyses, for example normalization of time series of nutrient inputs and trend analysis of the resulting time series, is greatly enhanced when conducting an initial analysis of the data quality. In general, data quality should be ensured by checking the data for gaps, i.e. missing values, and for suspect values, i.e. outliers. When investigating suspect values, the data should be checked for analytical errors or errors in data storing process, for consistency with previously reported data and with data from other comparable sources, and for errors when transferring data between databases.

A first task in the establishment of a data quality routine is precise identification of gaps in the dataset (which variables are missing and what is the length of the missing period?), followed by determination of the type of gap (not measured, measured but not reported, etc.). Data gaps in time series on nutrient input may occur for a number of different reasons:

- Measurements are missing from a sub-catchment for certain periods of time.
- Measurements of nutrient concentrations are missing.
- Runoff has not been measured.
- Nutrient and runoff data are both missing for a certain period of time.
- Measurements could not be made due to external conditions (e.g. ice cover).
- Data have not been reported for unknown reason.
- Concentrations and/or runoff value seem suspect and have therefore been omitted from the calculation of inputs; alternative inputs have not been estimated.

Several different methods are available for filling in data gaps. Depending on type, any of the following methods can be applied to fill in the gap:

- The mean value of a statistical distribution. The distribution is determined either by including all relevant data on the given catchment or from a shorter time series, for instance when estimating missing data from point sources in the beginning or end of a time series.

- The mean of adjacent values. If x_a and x_c are perceived as two time series values with x_b missing, then:

$$x_b = \frac{x_a + x_c}{2} \qquad (2.1)$$

- Linear interpolation. If x_a and x_b are perceived as two adjacent values to n missing values, then the k^{th} missing value (from x_a) can be estimated as:

$$x_k = x_a + k \cdot \frac{x_b - x_a}{n+1} \qquad (2.2)$$

- If runoff is known and a good relationship can be established between nutrient input and runoff, this can be used to estimate missing values,

- A q-q relationship can be used to estimate missing runoff values; a good q-q relationship can often be established for a nearby river.

- A load-load relationship for another river for which high correlation can be verified.

- Model estimations of unmeasured catchment loads, if possible – otherwise, inputs can be estimated from reference data.

- Assignment of a real value in the interval between zero and the limit of detection (LOD)/limit of quantification (LOQ) to observations below a limit of detection/limit of quantification. The PLC guidelines (see chapter 5) describe how to handle concentrations under LOD/LOQ when calculating loads.

Most methods for trend analysis, like the Mann-Kendall's trend method (see chapter 5), can handle missing values, preferably in the middle and not at the end of the time series (e.g. either the first two or the last two years). The trend test will be only negligibly affected if missing values are few. The statistical power of the trend tests decreases if the time series show gaps as it is more difficult to prove a real trend significant at reduced statistical power. If many missing values have been estimated and the inserted values are the same for many years, a trend test should not be performed as variation will be much smaller than when the data are based on real observations.

Above, various methods for filling in gaps have been described. Usually, the circumstance will decide which method to choose, but the following rank exists:

1. A model approach – i.e. a regression type model – to estimate nutrient load or flow.
2. Linear interpolation.
3. Values from a look-up table or values provided by experts.
4. No filling in of gaps. The time series is used as it is and assessments are made afterwards.

Outliers are data values that are extreme compared to other reported values for the same locality (country, basin, catchment, etc.) and can only be determined and flagged by conducting a formal outlier test using for instance:

- Dixon's 4 sigma (σ) test: Outliers are the values outside the interval consisting of the mean ±4 times the standard deviation.
- A box and whisker diagram.
- Experience-based definition of maximum and minimum values that is not likely to be exceeded or fallen below.
- Water quality standards (interval values or limits), if available.

It is important to note that outliers are not necessarily faulty data, but data requiring extra careful evaluation prior to use in statistical analyses.

Suspect or dubious values are values that do not fulfill the requirement of being determined as a formal outlier but differ significantly from the remaining values in the time series, or values that are unreliable; for instance, a load value for the reported runoff or data from a neighboring catchment. Suspect or dubious values may occur if measurements in a sub-catchment have been made for only a limited period of time, if changes in laboratory standards have occurred, or if changes have been made in other measurement methods, resulting in an abrupt change in data values. Also calculation mistakes may occur due to use of wrong units, faulty water samples, laboratory mistakes, etc. Suspect or dubious values should be corrected and treated as a formal outlier unless they can be proven correct.

If a dubious value is determined, deemed to be wrong and omitted from assessments, and if it is not possible for the Contracting Party to correct the value, it should be removed from the PLC database by the Contracting Party. If a reported data value is determined to be an outlier and deemed to be omitted from assessments, the outlier can be replaced in the assessment using a method from the list on data gaps. Usually, filling in data gaps or replacing suspect data cannot substitute measured data; thus, if possible, preferably measured or consistent model data should be found and used. It should be stressed that filled-in data gaps must be clearly marked in the PLC database.

3 Uncertainty of inputs (yearly input from a specific country or area)

Time series of nutrient inputs demonstrate a certain amount of year-to-year variation due to the contributions from a large number of different components. One such component is a possible trend in inputs over time, and time series are therefore, by standard, detrended before analysis of variance components since trend-induced variations are not of basic interest in estimating the variance components.

In the case of a time series with a constant mean value, i.e. no trend present, the time series will either be detrended or tested to avoid a significant upward or downward trend. Variation appears within the yearly values – and it is thus assumed that the yearly inputs are sampled from the same population of inputs with a given mean value and a given variation. This variation is, in fact, an estimate of the total uncertainty of a given yearly input, i.e. the standard error of the mean.

Total uncertainty is a complex sum (based on certain assumptions) of a number of different uncertainty components:

- Uncertainty due to field sampling (uncertainty from field sampling/measurements of concentrations of nutrients, metals and other substances, uncertainty from measurements of water velocity and stage, etc.).
- Laboratory uncertainty (variations in components lend uncertainty to laboratory analysis processes).
- Uncertainty deriving from the sampling set-up (how often, where and when, sampling location, time) and the methods for calculating runoff (either stage-discharge relationship or other methods) and load (based on combined concentrations and runoff).
- Variation introduced by year-to-year differences in climate (amount, type, and distribution of rainfall and changes in accumulated pools (snow/ice, soil and groundwater)).
- Uncertainty from estimation of unmeasured loads (bias from omitting unmeasured loads and uncertainty of the methods applied for estimating unmonitored loads).
- Uncertainty of inputs from direct point sources, including sampling, analytical errors, etc.
- Most probably, several other components contributing to uncertainty.

Awareness exists in most countries of analysis (laboratory) uncertainty, at least regarding nutrients. This is relatively well documented but may be one of the components contributing the least to total uncertainty. Most other components are complex, and some of them are very difficult to estimate in practice due to unavailability of empirical data. Uncertainty can be diminished by optimizing, for instance, time and location of sampling and implementation of a monitoring program taking into account variations in concentrations and runoff. An optimized monitoring program may introduce more strategic monitoring and more precise and modern techniques as well as an optimized methodology for estimating loads from unmonitored areas, strategic measuring being most important factor to decrease uncertainty.

Knowing the size of the different uncertainty components is not necessary when - as will be discussed later – testing for trends and for compliance with set targets. Variance component analysis is used in statistics when the researcher seeks to optimize the sampling design in a hierarchical sampling regime and/or to test for effects (treatment, emission reducing measures or other factors) using the correct sums of squares.

In the PLC-6 assessment, it would be useful to compare the total uncertainty of detrended nutrient load time series among countries, among sub-basins, etc., to determine if time series have the same level of uncertainty or if some countries, sub-basins, etc., have significantly lower or higher uncertainties. Investigation of the size of the different variance components would be highly useful for determining the reasons for the differences. The main result of such an exercise would be an overall improved data quality with more complete and consistent data sets from all Contracting Parties.

For this purpose, we need a standardised methodology for estimating the uncertainties in the national datasets. One such methodology for estimating the uncertainty of data from monitored rivers has been described in a paper by Harmel et al. (2009). The method is called DUET-H/WQ (software is available at the HELCOM web pages), which is based on the so called RMSE (root mean square error) propagation method. It is a fair approximation to the true value, which is often very complicated to derive.

In DUET-H/WQ, the uncertainty of individual measured loads is estimated by the formula:

$$EP = \sqrt{E_Q^2 + E_C^2 + E_{PS}^2 + E_A^2 + E_{DPM}^2}, \qquad (3.1)$$

where according to Harmel et al. (2009):

E_Q=Uncertainty of the discharge measurement (±%)

E_C=Uncertainty of sample collection (±%)

E_{PS}=Uncertainty of sample preservation/storage (±%)

E_A=Uncertainty of laboratory analysis (±%)

E_{DPM}=Uncertainty of data processing and data management (±%), i.e. load calculation or model uncertainty (see Silgram and Schoumans (ed., 2004)).

Then, the total uncertainty for aggregated data can be estimated by the formula:

$$EP_{total} = \frac{100}{\sum_{i=1}^{n} x_i} \sqrt{\sum_{i=1}^{n} \left(x_i \cdot \frac{EP_i}{100}\right)^2} \qquad (3.2)$$

and EP_{total} is given as ±%. EP_{total} is the uncertainty for the sum $x = \sum_{i=1}^{n} x_i$, where x_i is the monthly load from a catchment or a country.

The Contracting Parties will need to gather information on the different uncertainties, either from empirical data or from national or international papers and reports based on the same kind of data, i.e. riverine measurements based on more or less similar methods.

Furthermore, uncertainties regarding input estimates from unmonitored areas need to be described in order to estimate the total uncertainty for the whole catchment area. Uncertainty on direct inputs can be estimated using the same formula as above.

As mentioned in the beginning of this chapter, total uncertainty may also be estimated from the variance of a time series of inputs without trends or a detrended time series. It is the standard error of the mean input throughout the period. The two estimates of total uncertainty can then be compared. Both of the methods described here did not detect a systematical measurement bias, i.e. in runoff or in phosphorus inputs. Rather, they estimated the variation around an average value.

In a situation where the given time series of inputs show a significant positive serial correlation, the standard error is underestimated and total uncertainty is accordingly underestimated. In this report, we assume that the serial correlation in a yearly time series of nutrient inputs is small; the basic calculation of the standard error is therefore used as a close approximation to the true value of the standard error.

The method by Harmel et al. (2009) is illustrated by the following two examples: 1) total uncertainty for a river with high measurement precision and 2) total uncertainty for a river with low measurement precision.

Variance components	Example 1	Example 2
E_Q	5%	50%
E_C	5%	100%
E_{PS}	5%	30%
E_A	5%	25%
E_{DPM}	5%	50%

In Example 1 EP is 11% and in Example 2 EP is 125% when using formula 3.1. Total uncertainty of assuming a constant monthly input of 2500 tons (x_i) is 3% for Example 1 and 36% for Example 2. Total uncertainties were calculated using formula 3.2.

4 Hydrological normalization of nutrient inputs

The annual riverine inputs of nutrients show large variations between the reported years. Variation in runoff is a major reason behind this and is mainly caused by climate effects on hydrological factors such as precipitation, including accumulation and melting of snow/ice, and evapotranspiration, but also by temperature, etc. To remove the main part of the variation introduced by hydrological factors, the annual nutrient inputs are flow-normalized. Care should be taken when normalizing data if point sources have a large impact on calculated inputs, especially during periods with low water flow. Normalization should therefore not be applied to input from point sources discharging directly to the sea.

Normalization of riverine inputs is a statistical method whose result is a new time series of nutrient inputs where the major part of the hydrology-introduced variation has been removed. The normalized time series has a reduced between-year variation and the trend analysis is thus much more precise. Significant trends in the normalized series can probably be attributed to an effect of human activities.

Different methods for normalizing inputs are described in Silgram and Schoumans (ed., 2004), chapter 4. In this report, we focus on methods based on empirical data. The empirical hydrological normalization method is based on the regression of annual loads and annual runoff; thus, the method normalizes the loads to an average runoff (averaged over the time series period). In this way, the variation attributable to the annual amount of runoff is removed, whereas the effect of differences in the distribution of runoff over the year is not removed. In Silgram and Schoumans (ed., 2004), the normalization is based on un-transformed loads and runoffs. In our experience, the regression explains slightly more of the variation if both annual input and annual runoff values are transformed by the natural logarithmic function before normalizing.

The hydrological normalization should be regarded as a prerequisite for analysing trends. The trend analysis is a two-step process including: 1) the normalization and 2) the actual trend analysis.

According to Silgram and Schoumans (ed., 2004), the empirical hydrological normalization method should be based on the linear relationship between annual runoff (Q) and the annual load (L) of a nutrient:

$$L_i = \alpha + \beta \cdot Q_i + \varepsilon_i, \qquad (4.1)$$

where α and β are parameters associated with linear regression, and ε_i stands for the residual error in the linear regression. Then, the normalized load is calculated as:

$$L_{iN} = L_i - (Q_i - \bar{Q}) \cdot \hat{\beta}, \qquad (4.2)$$

where $\bar{Q}$ is the average runoff for the whole time series period. To avoid possible negative loads, the below formula should be used:

$$L_{iN} = L_i \cdot \frac{\hat{\alpha}+\hat{\beta}\cdot\bar{Q}}{\hat{\alpha}+\hat{\beta}\cdot Q_i}. \qquad (4.3)$$

Normally, the relationship is modelled after log-log transformation, reducing the influence of large loads and runoff values giving a slightly more precise fit with residuals that are more likely to be Gaussian distributed, which is a statistical prerequisite for the regression method. Thus, normalization should be based on a log-log regression between load and runoff:

$$\log L_i = \alpha + \beta \cdot \log Q_i + \varepsilon_i. \qquad (4.4)$$

This gives the following formula for normalized loads:

$$L_{iN} = \exp\left(\log L_i - \left(\log Q_i - \log \bar{Q}\right) \cdot \hat{\beta}\right) \cdot \exp(0.5 \cdot \mathrm{MSE}), \qquad (4.5)$$

and the following to avoid negative loads:

$$L_{iN} = \exp\left(\log L_i \cdot \frac{\hat{\alpha}+\hat{\beta}\cdot\log\bar{Q}}{\hat{\alpha}+\hat{\beta}\cdot\log Q_i}\right) \cdot \exp(0.5 \cdot \mathrm{MSE}). \qquad (4.6)$$

In the above formula (4.6), "log" is the natural logarithmic function, "exp" is the exponential function, and MSE stands for Mean Squared Error and is derived by the regression analysis (Snedecor and Cochran, 1989). MSE is calculated in all standard statistical software programs and is defined as:

$$\mathrm{MSE}=\frac{1}{n-2}\sum_{i=1}^{n}(x_i - \hat{x_i})^2,$$

where n is the number of observations in the time series, x_i is the observed value, and $\hat{x_i}$ is the modeled value from linear regression.

The factor "exp(0.5·MSE)" in the formulae is a bias correction factor and is derived as described by Ferguson (1986). The factor is needed in order to back-transform to a mean value and not to a geometric mean whose calculation does not require this factor. The main reason for using the natural logarithmic function for transformation is stabilization of the variance among residuals. Without the transformation, residuals are often distributed with a heavy tail to the right. Formula (4.6) is the recommended method for PLC-5.5 and onwards.

In PLC-5, the following method was used:

$$\log_{10} L_{iN} = \log_{10} L_i \cdot \frac{\hat{\alpha}+\hat{\beta}\cdot\log_{10}\bar{Q}}{\hat{\alpha}+\hat{\beta}\cdot\log_{10} Q_i}, \qquad (4.7)$$

after which the power function was used to back-transform formula 4.7. This method gives normalized loads which are a bit too low. Use of the natural logarithmic function has a more solid foundation in statistics than the base 10 logarithmic function. In principle, the presented methods can be applied even with a significant trend in the runoff time series, as long as the relationship between runoff and load is unchanged. Usually, the relationship changes with a significant change in the amount of runoff over time. This implies that a trend analysis of the runoff time series is needed in order to determine whether an upward or downward trend in the flow is present. If a trend occurs, we refer to Silgram and Schoumans (ed., 2004) for a method for normalizing loads.

In general, the differences between the methods are small, but especially for time series with a large year-to-year variation, methods without a correction term will give biased values with an underestimation of the normalized loads. This can have an unwanted effect when testing fulfillment of targets.

Hydrological normalization should be carried out catchment-wise, i.e. nutrient loads should be normalized for each catchment separately. If the normalization is performed country-wise or sub-basin-wise, the result will not be the same as the catchment-wise normalized nutrient loads summed to country or sub-basin level.

To illustrate the method, we used data from the Vistula River, Poland, to normalize both the load of total nitrogen and total phosphorus. Figure 4.1 shows scatter plots and the linear relation between loads and flow. Figure 4.2 shows the normalized time series together with the unnormalized loads. Note the large reduction in between-year variation in the normalized time series.

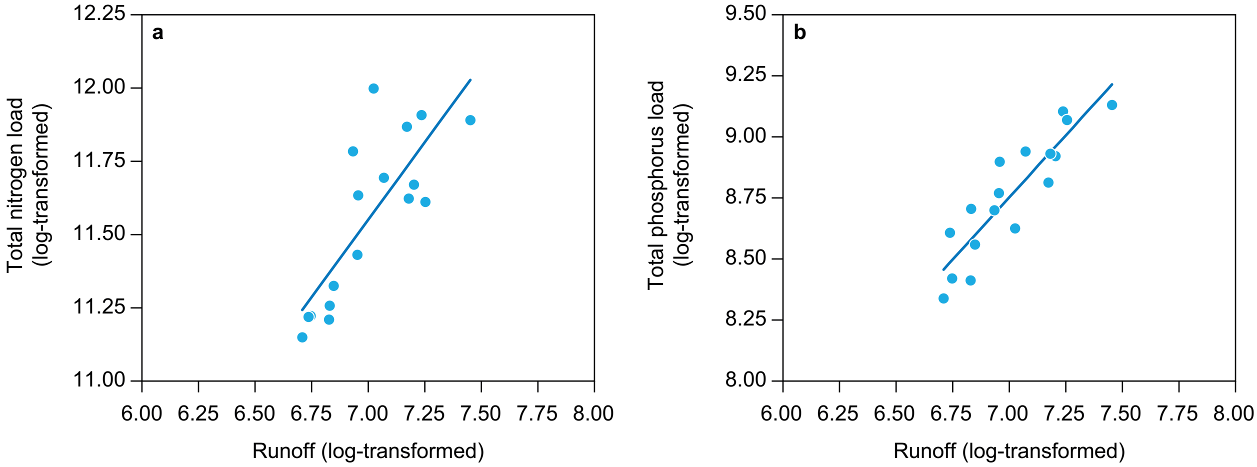
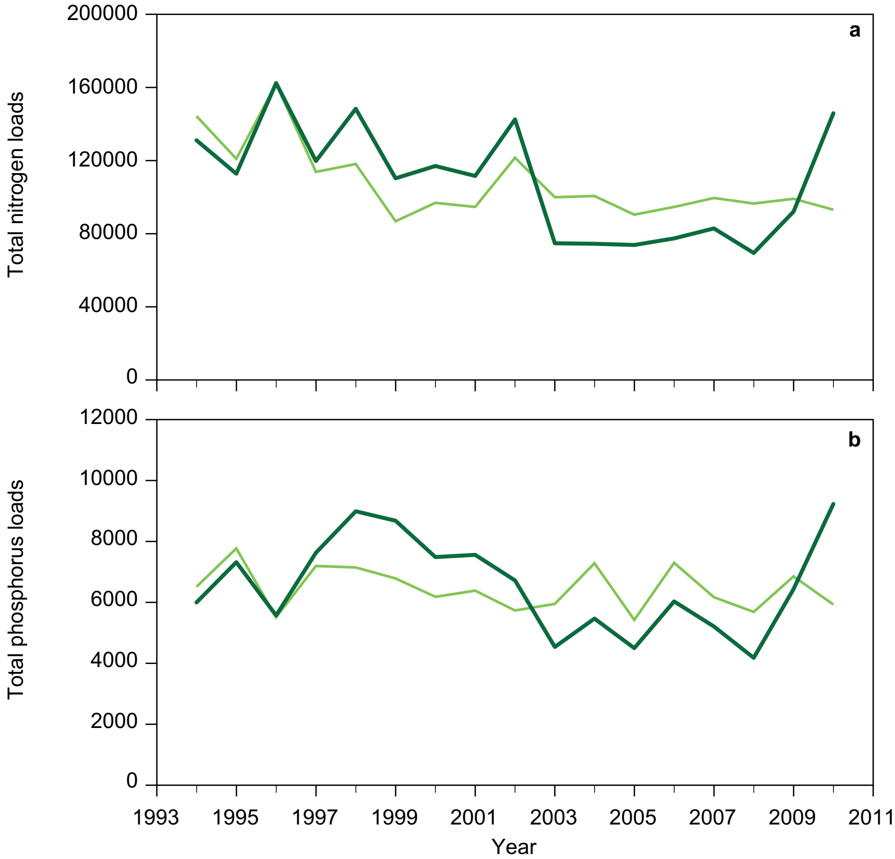


Figure 4.1. Scatter plots of annual loads of total nitrogen (a) and phosphorus (b) against runoff. Data represent the load of nitrogen and phosphorus to the Baltic Sea from the Vistula River in Poland during 1994-2010.

Figure 4.2. Time series plot of "raw" time series and of normalized time series of annual nitrogen (a) and phosphorus (b). Normalized time series are the green lines. Data are from the Vistula River in Poland.



5 Trend analysis and estimation of change

An important task in the PLC-6 assessment and the development of HELCOM CORE input pressure indicators is to perform trend analysis on normalized time series of nutrient inputs to different parts of the Baltic Sea, including trend analysis of the water runoff, with the purpose of evaluating if nutrient inputs are reduced and supporting evaluations of the effects of implemented measures and determination of whether the country-allocated reduction targets are fulfilled. The time series used in the trend analysis should always be normalized, but the methods described below may, of course, be used to analyse trends in unnormalised nutrient inputs as well. Trend analysis can be performed using a range of different both parametric and non-parametric methods. Parametric methods comprise ordinary regression with year as the independent variable and linear and non-linear regression methods, such as polynomial, exponential or more complex regression methods. The most well-known non-parametric method is the Mann-Kendall trend test and the Theil-Sen estimator for the yearly change in nutrient input. Apart from describing trend analysis methods, we will in this chapter treat methods for estimating the size of the trend when it is not linear.

The Mann-Kendall method (Hirsch et al., 1982) is a well-established method for testing for a monotone trend in a time series. It is non-parametric and based on Kendall's tau, which is a measure of the correlation between two different variables. The method is robust towards outliers and a few missing data. If the trend is linear, Mann-Kendall's method has slightly less power than ordinary regression analysis. The Annex gives a detailed mathematical description of the method, and software can be downloaded for free at http://en.ilmatieteenlaitos.fi/makesens or http://www.miljostatistik.se.

Ordinary regression analysis is also a well-known statistical method, but demands a linear relationship with Gaussian distributed residuals, which are stochastic independent as well (Snedecor and Cochran, 1989). If the time series is serially correlated, both the Mann-Kendall test and ordinary regression must be modified, since the tests will be impacted by this, and the probabilities of statistical test values can therefore not be trusted. On the other hand, it appears that the autocorrelation for annual time series of either loads or runoff is small and can be ignored; thus, the methods can be used without modifications as a good approximation. The minimum time series length for application of the Mann-Kendall test is 5 years. Accordingly, the trend analysis method allows itself to be more easily standardised and not slightly modified for each different time series, as is the case in the event of autocorrelation.

Both Mann-Kendall's trend analysis and ordinary linear regression allow performance of a one-sided trend test if focus is on testing for a downward or increasing development in a time series. This is of relevance in the development of the PLC-6, Core Input pressure indicators.

If a time series plot shows a clear trend reversal (also called a change-point in time), i.e. when the first part of the time series shows a linear increase and the second part shows a linear decrease in nutrient inputs, the analysis can be carried out by using a model with two linear curves ("the 2 sections method") or by applying two Mann-Kendall trend tests if both time series include a sufficient number of years (example in figure 5.2). Year of trend reversal (the change-point) can either be determined by inspecting the time series plot or by applying a statistical method (Carstensen and Larsen, 2006). If an exact year of change in the inputs is known (changes to sewage plants, etc.), this year should, of course, be applied as change-point, and the time series should be divided accordingly. Statistical estimation of the time when a change occurs in a time series is complex and involves a calculation procedure with iterative estimations. It is therefore suggested to determine the change-point by visually inspecting the time series plot if the time of change is not known prior to analysis. Time series analysis including more than 2 sections may occur, but "2 sections analysis" is a method better suited for application to very long time series, i.e. longer than the length of the PLC-6 time series.

The second part of trend analysis is the task of estimating the size of the trend or the change per year. Again, several different methods exist, and the specific use of these depends on the shape of the trend. The Theil-Sen slope estimator (Hirsch et al., 1982) is a non-parametric estimator that is resistant towards outliers (suspect) values. The method assumes a linear trend and estimates the change per year, and the estimator fails if the trend is non-linear, and if the time series shows time reversal, it is necessary two split the time series into two parts.

The size of a linear trend can also be estimated by regression. This is the classical approach, which is, however, not flexible with regard to all shapes of trend. The simplest method is using the start and end values in the time series of flow-normalized inputs, but if start and/or end values are too distant from the general trend, this method is not reliable.

If we seek to identify the total change in nutrient inputs over the whole time series expressed as a percentage, we can use the two methods below. Estimated linear slope:

$$100 \cdot \frac{(n-1)\cdot\hat{\beta}}{\hat{\alpha}}, \qquad (5.1)$$

where n is the length of the series, $\hat{\alpha}$ is the estimated input at start year minus one year, and $\hat{\beta}$ is the estimated slope. Formula 5.1 is based on the Theil-Sen slope estimator, and α is estimated using the estimator suggested by Conover (1980). When using start and end values we have the formula:

$$100 \cdot (\text{end-start})/\text{start}. \qquad (5.2)$$

For some times series, the start value, the end value or both can deviate too much from the general trend; if so, an approach using the average value of, for instance, the first 3 years and the last 3 years would reduce the influence of single years.

The trend analysis methods are illustrated below based on the time series of normalized total nitrogen and normalized total phosphorus inputs to the Baltic Sea from Poland. Contributions from direct point sources have not been added to the inputs. In figure 5.1, the normalized time series are shown together with a linear fit of the trends. A trend analysis should always be initiated with a time series plot of the data series. Table 5.1 includes the results of the trend analysis.

Figure 5.1. The fitting of linear regressions to flow-normalized inputs of nitrogen (a) and phosphorus (b). Data are riverine inputs from Poland.

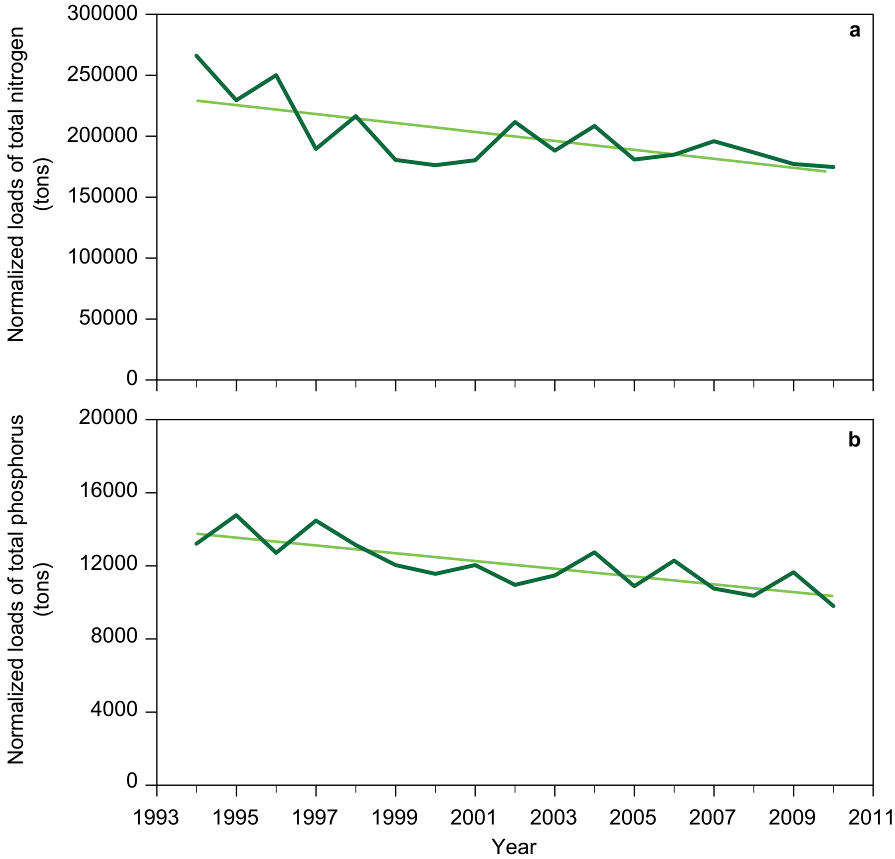


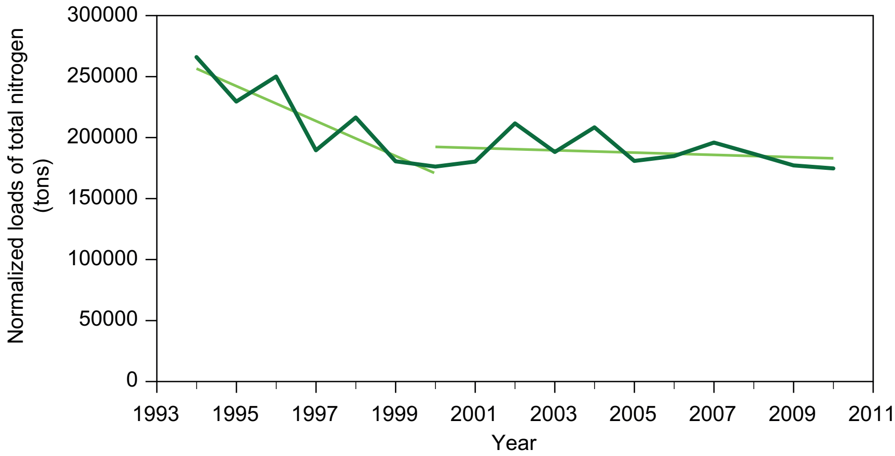
Table 5.1. Results of trend analysis of riverine inputs from Poland. Trends are significant when $P < 0.05$.

Time series	Runoff	TN raw	TN norm	TP raw	TP norm
Mann-Kendall	β=-0.724	β=-6581	β=-3525	β=-370	β=-0.215
	P=0.48	P=0.036	P=0.0074	P=0.0435	P=0.0003
Regression	β=-0.455	β=-5400	β=-3672	β=--330	β=-213
	P=0.46	P=0.023	P=0.0024	P=0.011	P=0.0002

The estimated change over the whole period for the normalized total nitrogen inputs is -26% according to formula 5.1 and -34% according to formula 5.2. The large difference in estimated change is due to the fact that the trend is not linear over the total period from 1994 to 2010. There is a change-point in the time series around 2000. The first period has a much steeper downward trend than the later period (figure 5.2). Application of Mann-Kendall's trend test to the two periods 1994-2000 and 2000-2010 shows that the downward trend in the first period is significant, whereas the trend in the last period is insignificant.

For total phosphorus, the estimated change over the whole period is -25% when using formula 5.1 and -26% when using formula 5.2.

Figure 5.2. Illustration of the two-line method in trend analysis. Data are normalized riverine inputs of total nitrogen from Poland.



6 Testing fulfilment of BSAP reduction targets

The progress in nutrient input reduction can be tested by two different methods: 1) trend analysis of time series of normalized nutrient inputs, as discussed in chapter 5; and 2) statistical analysis of whether the country-wise nutrient reduction targets under BSAP have been significantly met by a Contracting Party. In this chapter, a statistical method for testing fulfillment of reduction targets is proposed, and a traffic light system is introduced to illustrate a country's progress towards fulfilling the targets. A statistical method for testing if a normalized nutrient time series has moved relative to a defined nutrient target is needed. For this purpose, a parametric method based on the simple test of the mean value in a sample of Gaussian distributed data is suggested – a method that is often referred to as the fail-safe principle.

Let us assume that we have a time series of normalized inputs. The time series is initially assumed to be without a statistical significant trend and without a significantly large serial correlation, and we assume that the reduction target T (or any kind of target such as, for instance, maximum country input target) is defined without error, i.e. is a fixed value (certain amount of nitrogen/phosphorus given without any uncertainty). Let us finally assume that the data is sampled from a Gaussian distribution with mean value μ and variance σ^2.

As null hypothesis for the statistical test, we assume that the target has not been fulfilled, i.e.:

$$H_0 : \mu \geq T ,$$

The alternative hypothesis $H_A : \mu < T$ follows from this, i.e. the target has been fulfilled. Now assume that the test probability α is defined to be 5% (0.05), and then calculate the statistic.

$$\bar{x}_{AD} = \bar{x} + 1.645 \cdot \mathrm{SE}, \qquad (6.1)$$

where $\bar{x}$ is the mean of all values in the time series and SE is the standard error (SE = standard deviation divided by square root of n = number of observations in the time series), and, finally, 1.645 is the 95% percentile in a Gaussian distribution with mean 0 and variance equal 1. A test probability of 5% means that we have a 5% probability of incorrectly rejecting the null hypothesis.

This statistic is called the adjusted mean, and if the statistic is less than the target T, the reduction target is fulfilled.

In the case of a time series on nutrient inputs with a significant trend, another statistical method is needed for testing if a BSAP target is fulfilled. Let us assume that the trend is linear, a linear regression model with year as independent variable can be fitted to the time series, estimates for a and β can be calculated, and the residuals are Gaussian distributed. The linear model is then used to predict a normalized nutrient input for the last year n in the time series. This estimate is calculated as:

$$\widehat{L_{nN}} = \hat{\alpha} + \hat{\beta} \cdot year_n. \qquad (6.2)$$

Next, we need the standard error of the prediction which is defined as:

$$\text{SEr}=\sqrt{MSE}\cdot\sqrt{1/n + {year_n}^2/\sum_{i=1}^{n} year_i^2} \qquad (6.3)$$

where MSE is the Mean Squared Error as defined in chapter 4, n is the number of years in the time series, $year_n$ is the last year in the time series (i.e. 2010), and $year_i$ simply stands for a given year in the time series (i.e. 1997). Then the statistic is calculated as:

$$\bar{x}_{AD} = \widehat{L_{nN}} + t_{n-2,0.05} \cdot \text{SE}r, \qquad (6.4)$$

where $t_{n-2,0.05}$ is the 95% percentile in a t-distribution with n-2 degrees of freedom. A list with the 95% percentiles for different values of n-2 is given in annex 2. The mathematical definition of the standard error of the prediction SE_r given in (6.3) is a well-known statistic from ordinary linear regression (Snedecor and Cochran, 1989). If the trend is not linear, another model has to be used for the time series, and the formula for the standard error needs to be revised. The form of the trend in the data will dictate the method to be applied.

If the final years in the time series differ substantially from earlier years, either in level or trend, the last portion of the time series should be used to perform the statistical evaluation of target fulfillment. The exact number of years to be included depends on the actual situation, but minimum 5 years should be included. If only a few of the last values in the time series show substantial differences, the full times series should be used.

Finally, a traffic light system can be defined to obtain a status of whether a country has met the set BSAP target, whether it is close to fulfilling the target, or whether the target has not been fulfilled. This is described in HELCOM LOAD 4/2012 doc 5/2.2. Statistically, we define the system:

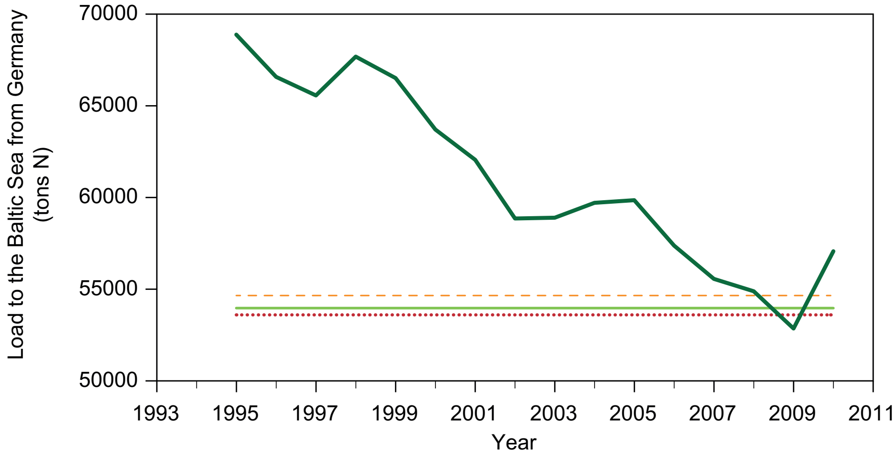
Red:
If $\bar{x}$ or $\widehat{L_{nN}} > T$, i.e. the average normalized nutrient input over the considered period or the estimated normalized input for the last year is above the target value.

Yellow:
If $\bar{x}$ or $\widehat{L_{nN}} < T$, and if $\bar{x}_{AD} > T$, i.e. the null hypothesis of target test is accepted, but the average normalized input or the estimated normalized input for the last year is lower than the target value.

Green:
If $\bar{x}_{AD} < T$, i.e. the null hypothesis of the target test is rejected.

To illustrate the principles, we tested if the total nitrogen inputs from Germany to the Baltic Sea (normalized riverine input + input from direct point sources + atmospheric deposition) met the provisional BSAP input ceiling (by August 2013) of 53,813 tons per year. A trend analysis of the summed up nitrogen inputs to the Baltic Sea from Germany showed a significant downward trend. In order to test target fulfillment, we fitted a linear regression to the time series and estimated the value in 2010 to be 53,590 tons. Using formula 6.4, the test value was 53,590+1.76*435=54,356, which is above the target. Traffic light evaluation resulted in a yellow light as the estimated value in 2010 was under the target value. The test principle is illustrated in figure 6.1.

Figure 6.1. Plot of principles for time series with trend created from German data on the total input of nitrogen to the Baltic Sea. Full line is target, "-----" line is estimated value in 2010, and "…." line is the test value according to formula 6.4.



7 Step by step analysis illustrated by HELCOM data examples

This chapter will present a full statistical analysis of time series from normalization in order to test whether a target has been fulfilled. We use data from the input of total waterborne phosphorus to the Kattegat (from both Sweden and Denmark).

We assume that the data have been evaluated for data gaps and outliers and thus are without missing values and errors – in other words, the data have been accepted by all relevant Contracting Parties.

The first hydrological normalization is performed for all rivers that discharge into the Kattegat from both Sweden and Denmark. The normalization of the River Göta Älv in Sweden is given as an example.

The relationship between log-transformed inputs of total phosphorus and runoff in River Göta Älv is shown in figure 7.1.

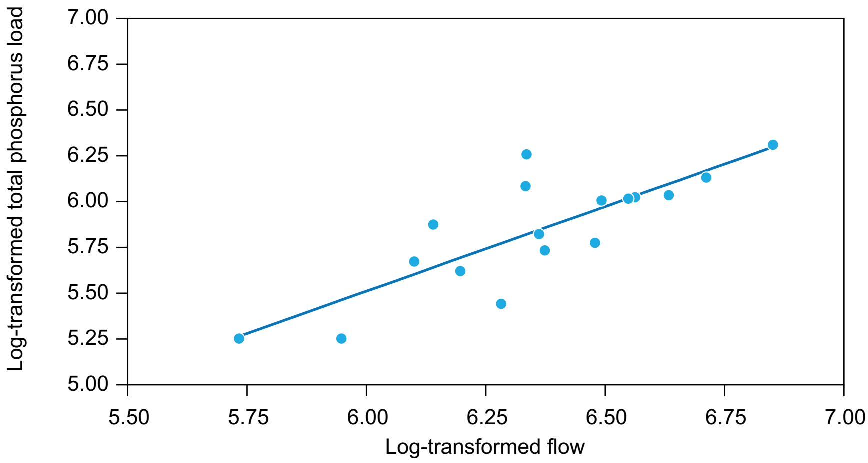


Figure 7.1. Linear regression on total waterborne phosphorus inputs and runoff for River Göta Älv in Sweden.

The next figure shows the normalized inputs for total phosphorus summed up for all rivers discharging into the Kattegat, plotted together with the measured unnormalized inputs. As can be seen, the variation between years is significantly reduced.

As mentioned, the normalization is carried out for all rivers discharging into the Kattegat, and these normalized inputs summed for all the rivers together with inputs from direct point sources and atmospheric deposition are used for the trend analysis and the target testing.

Figure 7.2. Normalized (green) and measured riverine inputs of total phosphorus to the Kattegat.

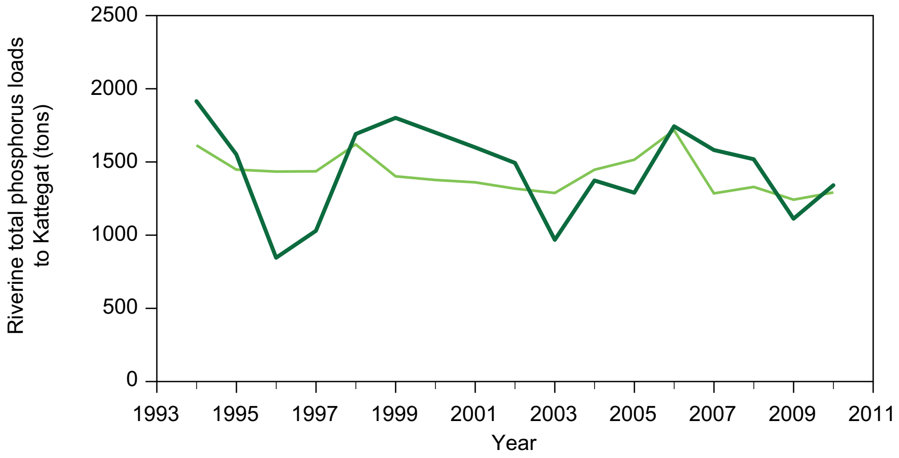
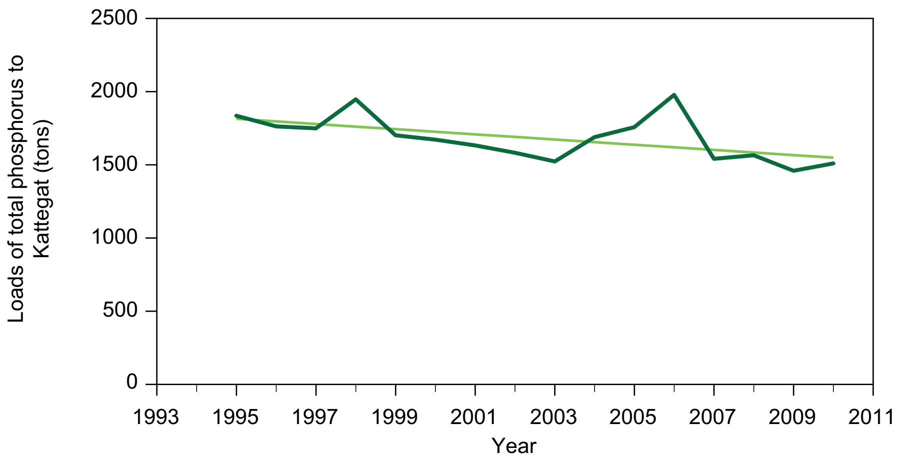


Figure 7.3 shows the linear trend line fitted through the time series of total phosphorus inputs. The trend in total phosphorus inputs to the Kattegat (water + airborne) seems to be close to linear.

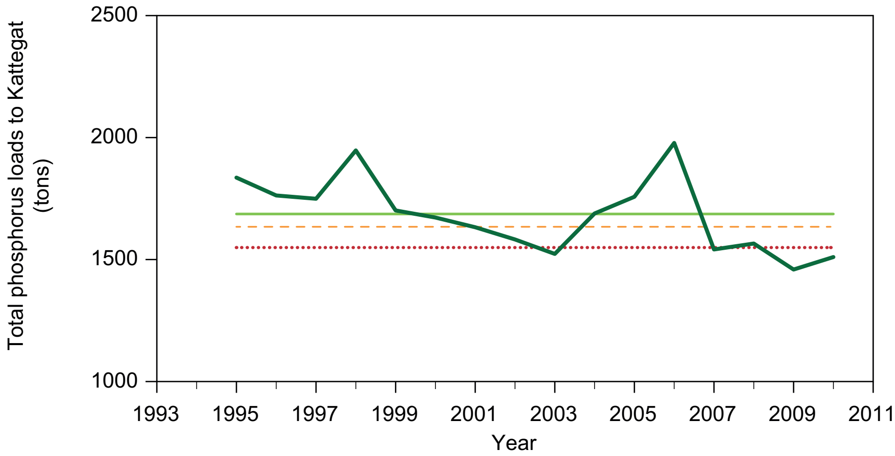
Figure 7.3. Linear trend fit to total water + airborne inputs to the Kattegat during 1995-2010.



The result of the Mann-Kendall trend test is a highly significant downward trend (two-side test, P=0.0060; one-side test, P=0.0030). The slope is estimated to be -21 tons per year (Theil-Sen slope). The total change in input over the period is estimated to be -17% (using formula 5.1).

The phosphorus input ceiling (target) for the Kattegat is set to 1,687 tons. The normalized total phosphorus inputs also show a significant downward trend, and when applying a linear trend to the time series, the normalized input in 2010 is estimated to 1,549 tons. Formula 6.4 gives the following test value: 1,549+1.76*47=1,634, which is 53 tons below the target. Therefore, a green light is given to inputs of phosphorus to the Kattegat.

Figure 7.4. Testing the target value for total water and airborne phosphorus to the Kattegat for the period 1995-2010. Full line is target, “------“ line is estimated value in 2010, and “…….” line is test value according to formula 6.4.



8 Discussion and recommendations

This report deals with the statistical aspects of analyses in relation to PLC data assessments, Core Input pressure indicator development, evaluation of fulfilment of BSAP reduction targets, etc. A number of different topics have been covered, for instance hydrological normalization, trend analysis, and significance tests for whether targets have been met or not. In the following, we have listed recommendations for which statistical method is best suited for the preparation of the PLC-6 guideline.

- Good data quality and consistency are a must to conduct reliable statistical analyses of the available time series. Time series may include gaps and/or suspect/dubious values. In chapter 2 of this report, methods for filling in gaps and how to determine if a dubious value is an outlier are described.

- Regarding total uncertainty in country data: It is a difficult task to calculate the exact uncertainty for the data provided by the contracting parties. One potential method may be to apply the simpler method DUET-H/WQ described in Harmel et al. (2009), which gives an approximation to the total uncertainty in monitored catchments. Information on the uncertainty of nutrient inputs in unmonitored areas has to be given by the Contracting Party – either by model uncertainty or as an expert evaluation.

- Normalization of nutrient inputs should be performed using the method based on transformed inputs and runoff. Transformation should be undertaken using the natural logarithmic function (see formula 4.6 in this report). Normalization is carried out for each catchment (river) separately, and normalized inputs can be summed up at country or at Baltic Sea sub-basin level. Normalization is a necessary step before conducting trend analysis. The method ensures that variation in annual inputs is significantly reduced, contributing to test for a significant trend in inputs by allowing identification of minor trends as being statistically significant. If a decision is made to use monthly input time series in the future, similar normalization methods can be applied to the monthly data (see Silgram and Schoumans (ed., 2004)).

- Concerning trend analysis, the Mann-Kendall non-parametric trend method is recommended for testing a significant monotone trend in the normalized time series. The method is fairly robust although autocorrelation can deflate the power of the test as it will for all statistical test methods. We assume that the autocorrelation in the yearly time series of nutrient inputs is of minor importance and therefore see the Mann-Kendall trend test as very good approximation. This non-parametric method can be used on both "raw" nutrient time series, normalized time series and runoff (climate) time series. If it is decided to use monthly input time series in the future, the Kendall trend test has been extended to a seasonal version (Hirsch and Slack, 1984).

- Estimating the change in nutrient inputs can be done by the non-parametric Theil-Sen slope estimator. The method assumes a constant change, i.e. a linear trend. Thus, use of the Theil-Sen slope estimator is recommended if the trend can be assumed to be fairly linear. If the trend is not linear, a non-linear model or start end difference should be used.

- If the time series show two or more distinct trends (trend reversal), two or more linear trends should be applied to model the time series. The change-point can either be determined by visual inspection of the time series plot or by a statistical method (Carstensen and Larsen, 2006).

- BSAP nutrient reduction target values have been defined, and a statistical method is needed in order to decide if the targets have been fulfilled. For time series with a non-significant trend, the equation in formula 6.1 can be used to calculate the adjusted mean nutrient input and evaluate this value against the target value. For time series with a significant linear trend, the equation in 6.4 should be used. We have defined a traffic light system allowing evaluation of nutrient inputs from varying catchments/Contracting Parties to the Baltic Sea according to defined targets.

9 References

BSAP 2007: HELCOM Baltic Sea Action Plan. HELCOM Ministerial Meeting, Krakow, Poland 15 November 2007,

Carstensen, J. and Larsen, S. E. (2006) Statistisk bearbejdning af overvågningsdata – Trendanalyser. NOVANA. (Statistical assessment of monitoring data. Trend analyses. NOVANA). Danmarks Miljøundersøgelser. 38 s. – Teknisk anvisning fra DMU nr. 24. http://www.dmu.dk/Pub/TA24.pdf (in Danish)

Conover, W.J. (1980). Practical Nonparametric Statistics. Second Edition. John Wiley and Sons, New York.

Ferguson, R. I., (1986) River loads underestimated by rating curves: Water Resources Research, 22, 1, p. 74-76.

Gilbert, R.O. (1987) Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold.

Harmel, R. D., Smith, D. R., King, K. W. and Slade, R. M. (2009) Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling application. Environmental Modeling & Software, 24, 832-842.

Hirsch, R.M. and Slack, J.R. (1984) A nonparametric trend test for seasonal data with serial dependence. Water Resources Research, 20, 727-732.

Hirsch, R.M., Slack, J.R. and Smith, R. A. (1982) Techniques of trend analysis for monthly water quality data. Water Resources Research, 18, 107-121.

Kendall, M.G. (1975) Rank Correlation Methods. Charles Griffin, London.

Silgram, M. and Schoumans, O. F. editors (2004) Modeling approaches: Model parameterization, calibration and performance assessment methods in the EUROHARP project. EUROHARP 8-2004.

Snedecor, G. W. and Cochran, G. W (1989) Statistical Methods. Iowa State University Press, Ames, Iowa.

HELCOM, 2013 by Svendsen, L.M., Staaf, H., Gustafsson, B., Pyhälä, M., Bartnicki, J., Knuuttila, S. nd Durkin, M. (2013). Review of the Fifth Baltic Sea Pollution Load Compiltion for the 2013 HELCOM Ministerial Meeting. Balt. Sea Environ. Proc. No. 141, 49 p

Annex 1: Mathematical description of the Mann-Kendall trend test

Trend analysis of a time series of length T and early loads of nutrients can be done by applying Mann-Kendall's trend test (Hirsch et al., 1982). This test method is also known as Kendall's τ (Kendall, 1975). The aim of this test is to show if a downward or upward trend over the period of T years is statistically significant, or if the time series merely consists of a set of random observations of a certain size. The Mann-Kendall's trend test has become a very effective and popular method for trend analysis of water quality data.

The Mann-Kendall's trend test is a non-parametric statistical method, which means that the method has fewer assumptions than a formal parametric test method. The data do not need to follow a Gaussian distribution as in ordinary linear regression but should be without serial correlation. Furthermore, the method tests for monotone trends and not necessarily linear trends, and it thus tests for a wider range of possible trend shapes. The direction of the monotone trends may be either downward or upward without any specific form. The power of the Kendall trend method is slightly lower than ordinary linear regression if the time series data are Gaussian distributed and the trend is actually linear, as this will encompass the slightly less restrictive assumptions.

Let $x_1, x_2, \ldots, x_n$ be yearly loads of total nitrogen or total phosphorus for the years $1, 2, \ldots, n$. The null hypothesis of the trend analysis is: the n yearly data values are randomly ordered. The null hypothesis is tested against the alternative hypothesis that the time series has a monotone trend. The Kendall statistic is calculated as (S = variance):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}\left(x_j - x_i\right),$$

where

$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

If either x_j or x_i is missing, then $\mathrm{sgn}\left(x_j - x_i\right) = 0$ per definition.

The trend is tested by calculating the test statistic:

$$Z = \begin{cases} \dfrac{S-1}{\left(\mathrm{var}(S)\right)^{\frac{1}{2}}} & S > 0 \\ 0 & S = 0 \\ \dfrac{S+1}{\left(\mathrm{var}(S)\right)^{\frac{1}{2}}} & S < 0 \end{cases}.$$

The variance S under the hypothesis of no trend is calculated as:

$$\mathrm{var}(S) = \frac{n(n-1)(2n+5)}{18},$$

where n is the number of loads in the time series.

A positive S-value indicates an upward trend and a negative value indicates a downward trend. When both a downward and an upward trend are of interest (a two-sided test), the null hypothesis of randomly ordered data is rejected when the numerical value of Z is less than the $\left(\alpha/2\right)$-percentile or greater than the $\left(1-\alpha/2\right)$-percentile (two-sided test) in the Gaussian distribution with mean value 0 and variance 1. A one-sided test can be carried out as well. The significance level α is typically 5%. The reason for evaluating Z in the standard Gaussian distribution is the fact that S under the null hypothesis is Gaussian distributed with mean value 0 and variance $\operatorname{var}(S)$ for $n \to \infty$. The Gaussian approximation is very good if $n \geq 10$, and fair for $5 \leq n \leq 10$.

It is possible to calculate an estimate of the trend β (a slope estimate) if one assumes that the trend is constant (linear) during the period and the estimate is change per year. Hirsch et al. (1982) introduced the Theil-Sen slope estimator, which can be calculated in the following way for all pair of observations $\left(x_i, x_j\right)$ with $1 \leq j < i \leq n$:

$$d_{ij} = \frac{x_i - x_j}{i - j}.$$

The slope estimator is the median value of all the d_{ij}-values and is a robust non-parametric estimator and will generally work for time series with serial correlation and non-Gaussian distributed data. A $100(1-\alpha)$ % confidence interval for the slope can be obtained by undertaking the below calculations (Gilbert, 1987).

Select the desired confidence level α (1, 5 or 10 %) and apply:

$$Z_{1-\alpha/2} = \begin{cases} 2{,}576 & \alpha = 0{,}01 \\ 1{,}960 & \alpha = 0{,}05 \\ 1{,}645 & \alpha = 0{,}10 \end{cases},$$

in the following calculations. It is standard to use a confidence level of 5%.

Calculate:

$$C_\alpha = Z_{1-\alpha/2} \cdot \left(\operatorname{var}(S)\right)^{\frac{1}{2}}.$$

Calculate:

$$M_1 = \frac{N - C_\alpha}{2},$$

and

$$M_2 = \frac{N + C_\alpha}{2},$$

where

$$N = \frac{1}{2} n(n-1).$$

Lower and upper confidence limits are the M_1th largest and the $(M_2 + 1)$th largest value of the *N* ranked slope estimates d_{ij} .

A non-parametric estimate for the intercept α can be calculated according to Conover (1980). The estimator is calculated as:

$$\hat{\alpha} = M_x - \hat{\beta} \cdot M_i,$$

where M_x is the median value of all the data in the time series, and M_i is the median value of $1, 2, \ldots, n$.

If the time series consists of data from different seasons (i.e. monthly loads), it is possible to apply Mann-Kendall's seasonal trend test (Hirsch and Slack, 1984). This is done by calculating the test statistic *S* for every season separately. Subsequently, the test statistic for the whole time series is equaled to the sum of each of the seasonal test statistics. We refer to Carstensen and Larsen (2006) for a detailed mathematical description of the seasonal trend test.

Annex 2: List of 95% percentiles of the t-distribution with n-2 degrees of freedom

***n*-2**	**95% percentile**
1	6.314
2	2.920
3	2.353
4	2.132
5	2.015
6	1.943
7	1.895
8	1.860
9	1.833
10	1.812
11	1.796
12	1.782
13	1.771
14	1.761
15	1.753
16	1.746
17	1.734
18	1.729
19	1.725
20	1.721
21	1.717
22	1.714
23	1.711
24	1.708

STATISTICAL ASPECTS IN RELATION TO BALTIC SEA POLLUTION LOAD COMPILATION

Task 1 under HELCOM PLC-6

HELCOM periodic pollution load compilation (PLC) assessments reports status and development in total annual runoff and total annual waterborne and airborne nutrient inputs to the Baltic Sea. This report deals with statistical methods for evaluating time series of annual runoff and nutrient inputs. Methods included are hydrological normalization of nutrient time series, trend analysis and a method for testing fulfilment of HELCOM Baltic Sea Action Plan (BSAP) nutrient reduction targets. Further is described how to fill in data gaps and to estimate the total uncertainty in nutrient inputs. These statistical methods are also included in the revised PLC guidelines.

ISBN: 978-87-7156-049-7
ISSN: 2245-019X