Document title	Statistical methodology report
Code	4-1
Category	INF
Agenda Item	4-Finalizing the PLC-7 products
Submission date	8.12.2021
Submitted by	Project Manager
Reference	

Background

PLC-8 IG-5 2021 took note of the status of the draft report in statistical methods uses for the PLC assessment. The meeting agreed that all comments should be provided to Denmark by 1 October 2021.

Denmark received comments from Sweden. The comments are for information included in this document (text with black color). With red color is DCE's answers on how the comments have been taken into in the final version of the report.

A nearly final version of the report was sent to PRESSRE 15-2021 for information.

The report is finalized and has been published on DCE homepage: https://dce2.au.dk/pub/TR224.pdf

During the finalization of the report based on the comments from Sweden and the internal quality procedures in DCE, some of the changes in the report required corresponding changes in chapter 10 in the PLC guidelines. These minor editorial changes were implemented in the version of the PLC guidelines that were adopted by PRESSURE 15-2021. Although, there are still pending some few editorial corrections needed in chapter 10 identified after the PRESSURE 15-2021.

Action requested

The Meeting is invited to take note of the finalization and publication of the statistical methodology report.

Comments to the technical report "Statistical aspects in relation to Baltic sea pollution load compilation" by Sweden

The following comments are made with focus on chapters 4 and 5 to identify paragraphs that might need further explanations or addition of information.

Chapter 4: Hydrological normalization of nutrient inputs

The problem of normalizing a time series that has a trend is discussed, traditional methods are presented and a method using differencing is suggested. The suggested method is illustrated by an example. The following questions arise:

(1) Would it not be better to model trend and normalization parameters in a combined model? It has been shown to give better results than doing the two modelling approaches stepwise, e.g. Smith and Rose (1991). Actually, it is the change over time in the relationship between loads and flows that we are dealing with. In multiple regression one has to remember, that the time variable is adjusted for the effect of the flow variable, so you normalize both the load and time with regard to the flow (see the paper by Smith and Rose). Therefore, the stage-wise and multiple regression will only give very different results if there is a trend in flow over time. Which of course is relevant to test. Which we also do and find only in a few rivers. In addition, it is difficult to understand the notion of an adjusted time according to flow. In addition, the flow-normalized loads are interesting data in their own right and important to report in the pollution load compilations and other Helcom reports.

(2) Changes in the concentration-discharge relationship over time might be observed, e.g. if the proportion between point and diffuse sources of pollution changes. While this might be important to optimally normalize data, the example shown in Figure 4.1. suggests that the changes are rather small, with two slopes being approximately equal and the third (red) one with a slightly lower slope. For the last line no discharge values below $7000 \cdot 10^6 \cdot m^3$ are observed which might influence the slope estimate. It is not obvious that this problem of differing slopes is serious enough to call for an adjusted method. There are no point sources included in the data presented in figure 4.1, actually I suggest not including point source data when you normalize, so only loads from diffuse sources should be normalized. It is correct that formally the slopes in figure are not significantly different in a statistical test, but looking at the residuals over time, you will see that they have a strong trend (positive to negative) with the former method whereas with the new method, no such trend in the residuals is present, which tells me that an adjustment is needed. Luckily, if the relationship between load and flow is not changing a lot over time, then the two methods give very close normalized values, so no harm done in using the new method for those as well, but the normalization is improved for those were the relationship changes over time.

(3) A method of differencing is suggested. This means that the change in transport from year t to year t+1 is modelled using the difference in discharge between the same two years. Formulas for the differencing are given in formula 4.7 to 4.9 and a normalization approach based on this is given in formula 4.10. While the model formula 4.9 includes a sum of all residuals for time points 1 to i, the normalization formula 4.19 seems to include only the estimated residual at time point i. It is unclear why. This is now made clearer in the text.

(4) The differencing method is illustrated by data for total phosphorus in river Neva. Using the traditional approach for normalization there is (almost) no relationship between discharge and total phosphorus. Looking at the differenced data in Figure 4.2 (second plot) there is some correlation, but probably weak as there is a lot of variation around the line. Single observations might have an influence on the slope of the fitted line. Probably there should be guidelines on when the fit to data can be assumed to be good enough to actually be used for normalization. In the current example, the estimated slope is used further seemingly without critical assessment on its value, which might have quite far-reaching consequences for the

produced normalized values. The NEVA example not really a good one and a new example is now presented in the text.

(5) Normalised values from the differencing method is presented in Figure 4.3, while it is obvious that the suggested method gives different results it is not clear that the results are better or more reliable than the results from the traditional method. It is clear from looking at the residuals of the model fitting. The model residuals are actually the only one you can use to distinguish between the different normalization methods. An additional figure showing this is now included.

Overall comment tom Chapter 4: The usefulness of the proposed model needs to be much better documented and referenced. A simulation study could be used to verify that the methods give reasonable results in realistic settings. A single and not obvious example is difficult to review. It should be shown why only the residual at time point i is used in computing the normalized values and not, as the formula indicates, all residuals up to i. Will do this at a later stage in time.

As for most series the relationship between loads and discharge is strong it is also clear that the traditional method works well in most cases and should be the standard case. An explicit trend estimate in the traditional normalization method is often used in practice and could improve the normalization when the time series is not stationary. In fact, an explicit trend estimate might lead to similar results as the suggested differencing method (Appendix A). As mentioned below, I prefer standard linear methods to GAM, because of a risk of overfitting with GAM, when used by researchers not familiar with GAM. Standard regressions are more know. If you read formula (4.9), you will see that, there is an explicit linear trend estimate there ($\alpha \cdot i$). But the error process is a AR(1) process instead of iid. That said, normally α can be accepted to be zero for most time series. That's because the term $\alpha \cdot I$ has more to do with a trend in the point sources,

To test and run the proposed normalization method, code is given in SAS. As SAS is an expensive commercial software that might not be accessible for HELCOM members it would be an advantage to give code also in a free software, e.g. R or Python. Will include coding for R.

5. Trend analysis, change point and estimation of change

Trends are suggested to be modelled using segmented regression, i.e. with change points. There are methods to automatically determine change points, but they need to be handled with care in series with few years and high variation. To handle this the report suggests that each part of the series should at least be five years, which is reasonable. It seems that the report suggests mainly to determine the change points from visual inspection, but this is not made clear enough. Now additionally, an iterative method is mentioned, and a reference is given. But if knowledge of possible change points in time are known from history regarding the catchment and regulations of pollution loads (point sources, farming regulations etc) this should of course be taken into account regarding the statistical analysis.

LOESS regression is shortly mentioned as alternative to segmented regression and should, in fact, lead to vary similar results, at least if variation is not too strong, as also shown in Figure 5.2. An advantage of LOESS is that it can faster adapt to changes, which is also seen in Figure 5.2, where the last seven years have constant values of total phosphorus around 4000, which is captured by LOESS, but not by the segmented regression. If prediction would be made from the two models the LOESS would probably lead to more reasonable results. Due to this, LOESS and similar methods such as generalized additive model, i.e splines fits for trends, could also be described in more detail and presented as valid alternative method. Uncertainty for such fits can be computed similarly as for traditional regression, even though there would be some new concepts to account for. E.g any uncertainty would be conditioned on the level of smoothing chosen for the model, but also this is similar to the currently suggested models, where the uncertainty is conditioned on the selected change points. Including LOESS and GAM would give a more complete picture of options available. This is now included as an option, but not described in detail, as I prefer the linear approach. Since overfitting is a risk with GAM's

Minor comments:

In formulas using the predicted normalised loads it would be more consistent with earlier formulas if the prediction are denoted $\widehat{L_{Nn}}$ and not$\widehat{L_{nN}}$ as the year is in other formulas set last. Yes, correct. It's changed now.

Different standard error are used in different chapters and it might be an advantage to use specification to show for which estimate the standard error is used, e.g., in formula 6.3 the standard error for the prediction for the last year is computed. As this is different from the standard error of other years it could be denoted $SE_{\widehat{L_{Nn}}}$. In formula 6.1 $SE_{\bar{x}}$ could be used to show the distinction. Yes, have inserted this.

For formula 6.5. the computed value is called control value while if before was called adjusted mean. In Figure 6.1 this value seems to be called test value. This makes that text sometimes difficult to read. Corrected.

Figure 5.4 could not be seen in the report. Yes, will be corrected for the final version.

The description of lines in the figure text of figure 6.1 and 7.3 is not obvious. A legend in the plot would simplify reading the plot. Will try to have legends inserted.

Annex 3: The following packages could be included for R (done)

Mann-Kendall test: package rkt

Nonlinear and loess regression: package gam and mgcv

Durbin-Watson: the function dwtest() is available in the lmtest package (i.e. it is not a package on its own), please clarify. durbinWatsonTest() is available in the car package.

References:

Smith, E., Rose, K., 1991. Trend Detection in the Presence of Covariates - Stagewise Versus Multiple-Regression. Environmetrics 2, 153–168. https://doi.org/10.1002/env.3770020204

Appendix A:

Code for a normalization with explicit smooth trend in the model. As correlations between discharge and load are weak any results must be handled with care.

Normalisation model with trend for Neva, q indicates log-transformed discharge, tp log-transformed total phosphorus:

```
model2<-gam(tp~q+s(year),data=neva)
```

Subtract discharge effects

```
neva$res<-neva$tp-model2$coefficients[2]*neva$q
```

Add mean discharge to bring series on the desired level.

```
neva$norm_trend<-neva$res+model2$coefficients[2]*mean(neva$q, na.rm=TRUE)
```

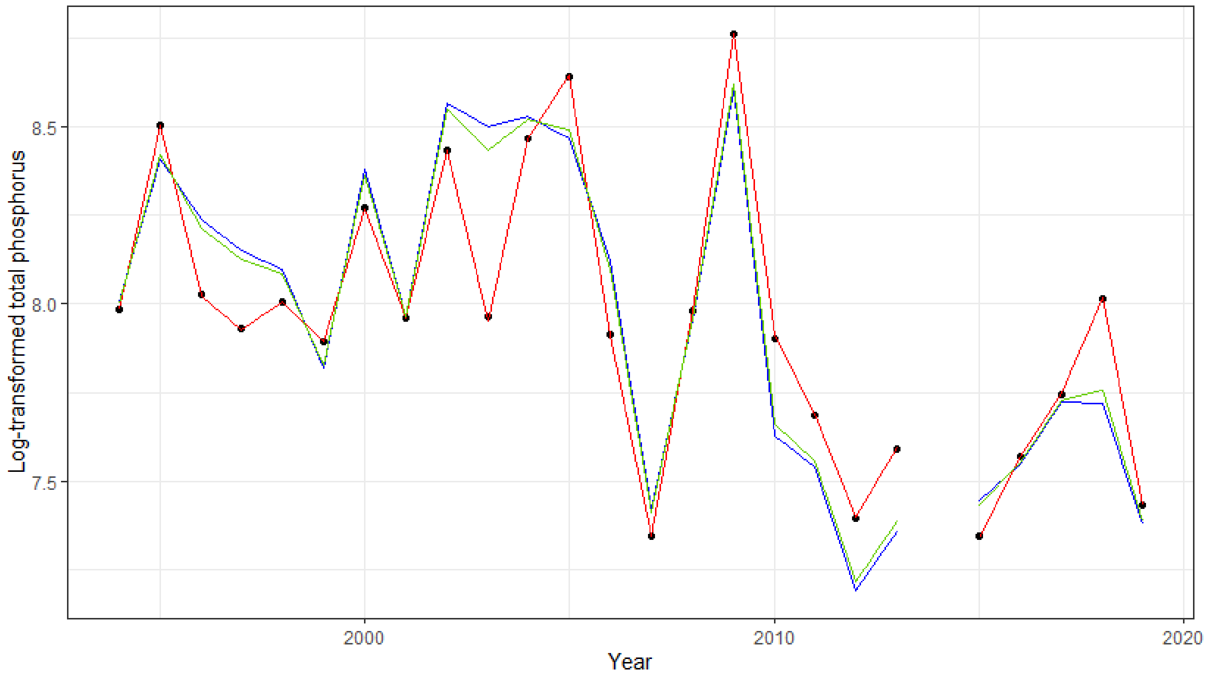

Figure A.1: Plot corresponding with Figure 4.3 in the report. Time series of annual actual values (black dots), the normalization method without explicit trend estimate (red), normalization using differencing (green) and normalization with an explicit trend estimate (blue).