Hydrology and Earth System Sciences Discussions

Interactive comment on "Statistical distribution of series of 12 monthly concentration samples for environmental classification of rivers" *by* J. Eliasson and T. Thordarson

J. Eliasson and T. Thordarson

Received and published: 4 January 2008

Anonymous Referee #3 Authors response included.

General comments

The paper is concerned with the important issue of water quality classification of rivers based on a limited amount of samples under economical constraints. To accomplish this task the authors investigate the distribution of monthly concentration values of different hydrochemical constituents of 14 rivers sampled over the period of one year. They compare it to the normal and lognormal distributions and propose a new cumulative distribution function (DoC) to allow water quality classification according to

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

Icelandic regulations. While the paper tries to tackle this interesting question, it has in my opinion some severe limitations:

1. I disagree with the authors assumption that 12-14 samples of different chemical parameters measured over a period of only one year are representative and would allow a sufficient classification and choice of the correct statistical distribution. This is especially the case when analysing the long tails of the distribution function as discussed in this paper. For a sufficient analysis and choice of the statistical distribution a much larger sample of concentration variables needs to be taken into account.

Response

Normally this is not possible, that is agreed. The problem is discussed on P2569 L10 and onwards. But the assumption is actually a conclusion drawn from the fact that the DoC pops up in 14 different rivers hundreds of kilometres apart as a result of a pooling process. We of course hope that the DoC will find a more general use in the future. A clarification of this point can be included at the end of the introduction chapter.

2. A further limitation is that the authors statistical analysis is based on the assumption that the concentration variables are independent from each other and identically distributed. This is the fundamental precondition of the central limit theorem where the authors refer to in the interpretation of Figure 5. But when dealing with the different chemical constituents of this study (e.g. pH, E.coli, TOC, IC, T-N etc.) one would expect correlations of some parameters that would violate this assumption. For instance the authors state that there is no strong seasonal correlation, but they ignore that some constituents are certainly correlated to the discharge. Neglecting the discharge conditions in the statistical analysis is a major constraint that needs to be addressed in this study.

Response

Identically distributed yes, not necessarily independent. 15 constituents are shown in

the figures, 11 are used in the pooling and 4 let out because of cross correlation as discussed on P2567 L15 so correlation is not totally ignored even though the exact numerical value of the coefficients are not used directly. It is interesting to note, however, that correlation does not destroy the distribution. Two random variables, say f and g where g = a + b f become one and the same when normalized, provided a and b are constants, pooling them produces the same value twice and the distribution is not affected but the IRL (Independent record length, 24 values instead of 12 in our case) and confidence limits are affected, see e.g. Buisand and Schaefer referred in Eliasson 1997.

3. The structure and language of the paper make it difficult for the reader to follow the line of thought. The objective and reasoning throughout the paper appear to be vague. For instance it does not become entirely clear how the major objective, the stream classification according to water quality, can be improved by the applied statistical methods. The paper would benefit in my view from a more precise objective and procedure in combination with a clear structure and revised language.

Response

Sorry, this has to be mended. Its to a degree inevitable as we try to use the normal distribution to start with. The DoC just pops up in the pooling, and a multitude of statistical methods can be applied to the data without ever finding it. But fig. 6 indicates its existence, if the 11 lines there had not been so close together that would be the end of our story, no DoC. In the next paper when we have more data the major objective of confirming or rejecting the DoC can be precisely stated. We can include a clarification on this point in a revised paper. This means a phrase change for clarification several places in the text.

Specific comments

1. The title of this paper may be misleading since 12 monthly concentration samples may imply 12 samples per month and not per year.

Response

Sorry, we thought this OK, our dictionary (The Shorter Oxford) gives monthly as Done or recurring once a month. We did not dream of that anyone should understand it otherwise.

2. P2562 L23 - P2564 L23: This column deals with the description of the study area and should be moved to section 2 Study area and sampling sites. Moreover a short introduction about the state of the art of water quality classifications based on a limited amount of samples and relations to other studies would be beneficial.

Response

This can be easily mended. The text discusses the following: Peculiarities of basaltic rock geology, results of Stefansson et al, influence of agriculture and the regulation system in Iceland. The relevant phrases can be taken out and included in Section 2.

3. P2566 Section 3.3: This section is named non-parametric approach although it deals with a parametric approach.

Response

What parameters ? This section discusses only ranked values of normalized data. Please clarify and we will mend it.

4. P2564 section 2: This section deals with the description of the sampling sites and sampling conditions. However the section lacks information about the general sampling scheme and sampling conditions, e.g. base flow sampling?

Response

Yes it does, both sampling techniques and analysis methods can be described in a much longer text. This is not done in the papers referred to, but the statistical results themselves emphasized instead. Is it mainly base flow sampling that is missing. If so it can be mended in aproximately half a page. Close description of general sampling

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

scheme, sampling conditions and analysis methods would take 10 pages or longer.

5. P2565 L15: The authors claim that the constituents do not have a strong seasonal correlation. Based on which statistical test was this conclusion drawn? Please specify strong.

Response

Correlation coefficients to air temperature are given in table 1. Seasonal variation is reflected in the air temperature in the northern hemisphere so geophysical variables with seasonal variation have either a high positive or negative correlation coefficient to air temperature. Water temperature (not a constituent) has a strong seasonal correlation (0.64 - 0.83). The constituents themselves have (numerically) much lower correlation to air temperature. There is no general agreement on the limit between strong and moderate correlation. If the DoC is accepted and used, we get more data and then a more rigorous definition for rejecting cross correlated series is needed. In the case presented in the paper, including the cross correlated series (4 out of 15) does not make any visual difference in fig. 6.

6. P2566 L4 onwards: The authors state that statistical tests do not reject the normal distribution for most of the series, although systematic derivation can be observed. Please specify which statistical tests have been made (e.g. Ks?) and for which series the test was rejected.

Response

We tried Shapiro-Wilks with confusing results. E. g. t-N got a reject in No 11 Botnsa but all the other constituents got an accept and t-N got an accept in all the other rivers. We consider the best test to be visual comparison of the lines in figs 2 and 3 to the red normal distribution line. The systematic deviations from the red line are evident. If they are statistically significant may be disputed. What we find indisputable is that the 11 distributions in fig. 6 are very much alike and there is no systematic deviation from the

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

red line (DoC). It is possible to include a TRUE and FALSE table for the Shapiro-Wilks (or other) test for normal distribution for the 15 constituents and 11 rivers, but we get a TRUE for many of the series with clear seasonal variation. We did not think this served a purpose.

7. P2572 L7 onwards: The authors argue that the pooling presumes that the constituents do belong to a common two parameter distribution, a presumption already in use in most environmental monitoring practice. As in this case, the authors frequently refer to the general practice without supporting their argumentation with references.

Response

There is a lot of textbooks in environmental statistics, they use the normal and the log normal distributions almost exclusively. These are two parameter distributions. Statistics of extremes use GEV which is the only three parameter distribution in general use that we know of.

8. The amount of references is quite limited and contains a lot of grey literature. Moreover two references of the annex are not mentioned in the document (ISO, State of New York Department of Health)

Response

Agreed. But that's what we could find and there are more blue than grey. Perhaps we are too stuck in the 12 monthly sample method. The ISO and New York come under no. 4.

9. It is not entirely clear what the correlation coefficient in Table 1 stands for.

Response

Its correlation to air temperature (Correl = 1). For use of correlation coefficients see no. 5.

Technical corrections

EGU

1. The display and layout of the Figures could be improved. In almost all Figures the differences between the constituents are hard to differentiate, a description of axis is lacking in Figure 6 and Figure 7 and axis devisions are sometimes overlaid by the graph like in Figure 7.

Response

Agreed. The reader actually has to read the paper online and use the possibility that the pdf offers, that is to enlarge the figures on the screen. We really have to congratulate the publisher on the very good resolution of the online paper, the small pictures in fig. 3 can actually be enlarged to fill out the screen without the lines getting blurred. Sorry about the flaws in figs 6 and 7, we hope that the reader forgives us. We thought that placing the vertical axes in the zero would actually be better than have it on the outside. The figures will be mended.

Acknowledgement We authors would like to extend our thanks to Anonymous Referee #3 for his valuable comments

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 4, 2561, 2007.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU