Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-39-RC2, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Detecting dominant changes in irregularly sampled multivariate water quality data sets" *by* Christian Lehr et al.

Anonymous Referee #2

Received and published: 22 March 2018

The manuscript proposes an exploratory framework for detection of dominant changes in multivariate water-quality data sets with irregular sampling in space and time. As stated in the introduction, many analysis methods assume regular temporal spacing, but many monitoring networks evolve over time resulting in irregularly spaced samples. The concept is good, some more effort needs to be put into the writing and analysis.

1. The abstract is rather lengthy.

2. The introduction contains vague statements and extraneous adverbs. The first sentence of the article is "Numerous high frequency studies unravelled the high temporal variability of stream water quality." This is well known, as shown by the many references. It seems like the first sentence of the article should start with a stronger sentence about the problem at hand. The second paragraph of the introduction has the

phrase "numerous different drivers at different scales." This is vague. Give an example, or qualify the drivers, such as climatic and land-use drivers. The second sentence of the third paragraph is either missing something or "determining" should be "determine."

3. In the description of the study area the mean annual precipitation and mean annual temperature are given for the federal state Brandenburg for 1961–1990. This does not overlap with the study period of 1990–2009 at all. With the common use and availability of climatic data, it would not take much effort to report precipitation and temperature for the study period. It is not clear what period the water balance variability values represent.

4. The topography and soils sections are well written and informative.

5. We know the data are collected irregularly, but are they collected to be representative of seasons and flow conditions, i.e., are there high-flow samples?

6. Figure 2 shows some sites with very little data, yet it seems like they were included. It is not clear how these help inform the method. It seems like there should be some minimum number of samples per year most of the years from 1998âĂŤ2009 in order for a site to be included in the study. Some parts of the proposed framework were done for sites with more than 50 observations. It seems like the entire analysis should be done only with those sites. It is not clear how these low-sample sites fit with the rest of the sites.

7. It has been very well documented that substituting a fraction of the reporting limit is an inappropriate method for dealing with censored data. See:

Gilliom, R.J., and Helsel, D.R., 1986, Estimation of distributional parameters for censored trace level water quality data, 1. Estimation techniques: Water Resources Reserach, 22, 135–146.

Singh, A., and Nocerino, J., 2002, Robust estimation of mean and variance using environmental data sets with below detection limit observations: Chemometrics and Intelli-

gent Laboratory Systems, 60, 69–86.

Helsel, D. R., 2005, More than obviousâĂŤBetter method better methods for interpreting nondetect data: Environ. Sci. Technol., 39(20), 419A–423A, DOI: 10.1021/es053368a

Helsel, D.R., 2005, Nondetects and Data Analysis: Wiley-Interscience, 250 p.

Helsel, D.R., 2006, Fabricating dataâĂŤHow substituting values for nondetects can ruin results, and what can be done about it: Chemosphere, 65(11), 2434–2439.

Helsel, D.R., 2012, Statistics for Censored Environmental Data Using Minitab and R: John Wiley & Sons, 324 p.

Admittedly, the percent of censored values is small, but substitution should really not be used anymore in water-quality analyses. I'm not sure if Isometric Feature Mapping can utilize censored values. However, the authors could estimate the mean and standard deviation of the constituents with censored values using regression on order statistics or maximum likelihood methods (see Helsel, 2012) before standardizing the variables. The Akritas-Thiel-Sen median line can be used for the trend analysis.

8. Check equation (2) in line 385. Should there be a plus sign between B0 and the summation symbol? Describe the components of the equation that were not already described in equation (1).

9. In the interpretation of components, the authors describe using multiple linear regression, which is a parametric method that assumes a model linear in the parameters, but then make an argument for a non-parametric measure of correlation applied to the multiple linear regression results. This seems contradictory.

10. Consider presenting the methods and the results in the same order for parallel construction.

11. In the discussion, the conclusions on page 32 about the 1st component were not

well supported. There were a lot of statements like "we assume a general effect," some process "might" happen, some processes "tend to enhance." The discussion of the 2nd component was better supported with information about the sediments in the area. Some of the material in the first paragraph of section 5.2 should be moved up to better support the conclusions about the 1st component. The discussion of the 4th component on page 33 seemed speculative. Has this been modelled or shown elsewhere?

12. Page 37 states nicely some important implications of the observed water quality.

13. Page 40, line 895, change "is" to "are."

14. Page 40, line 901, "Complementary" does not seem like an appropriate word for this sentence.

15. Some of the results, discussion, and conclusions mention both PCA and Isomap, but some of the numbers, figures, results must come from one of them specifically âĂŤthat should be made more clear.

16. Check that numbers in the text agree with the numbers in the figures and tables.

17. In suggesting this approach, how do you know the results are sufficient? Are there some measures of quality that can be incorporated into this?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-39, 2018.