Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-39-RC1, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

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

Anonymous Referee #1

Received and published: 16 March 2018

This manuscript presents a new exploratory framework for detection of dominant changes in multivariate water quality data sets with irregular sampling in space and time. The paper is well written and I think it is a valuable contribution to the hydrological community. I recommend its publication after the following comments are addressed.

General comments:

1. On the novelty of the proposed framework: I think this manuscript can foster future research ideas and efforts that are aimed toward detecting dominant changes in watershed using multivariate data at multiple sites. I think this type of coherent and systematic investigation of watershed data is limited in the literature, since previous studies have tended to focus on either only a few sites or a few constituents.

2. On the abstract: I found it quite lengthy (469 words), which prevents readers from quickly grasping the key messages. Also, it is not customary to have more than one paragraph in the abstract.

3. On the coverage of the monitoring data: The paper addresses the 'time' aspect of the collected water quality data but lacks a thorough discussion on the 'discharge' and 'season' aspects of those data. Were all constituents at these sites sampled roughly similarly across season? Were they sampled roughly similarly during normal-flow and stormflow conditions? Such information is important and can be simply shown with boxplots (e.g., with "month" and "discharge percentiles" as x-axes respectively.) If samples at these sites were not taken roughly similarly across season or discharge, how would that affect the validity of the proposed exploratory framework and the interpretation of the results? The authors should comment on that.

4. On the general applicability of the framework: Several points shall be discussed by the authors regarding the applicability of the framework, which can guide its application to monitoring network elsewhere. a) Is the framework intended to solute data only? Sediment and total phosphorus are typically monitored by many programs. Do the authors recommend the inclusion of such constituents in the proposed framework? b) What is the threshold for a constituent (or a site) to be included in the analysis? Specifically, how many samples are required for a constituent-site pair to be eligible? I am puzzled by the few stations in Figure 2 that have only 1-8 samples. I wonder whether these site-constituent pairs should be disregarded. c) For such multi-site and multi-constituent exploration, all available data should be considered to enhance the robustness of modeling results. However, not all the data are consistently available across the sites. Then, how should one handle the tradeoff between the number of constituents and the number of sites? If we rank all constituents by the number of applicable sites, C1, C2, C3, C3, ..., C16, then what is the relative gain of sequentially adding extra constituents (from C1 to C16) into the analysis framework? Can an explicit rule be developed to prevent adding new constituents to the framework?

5. On the irregularity nature of the monitoring data: The authors have provided adequate references in many parts of the manuscript. One exception is on the irregularity of water quality data ($\sim$ line 110 and also Section 5.4). One reference that you may find useful is provided below, which discusses at least two points that are discussed in this manuscript, including (a) irregularity nature of water quality data and how to model that property and (b) fractal scaling in water quality data which may affect trend significance (including the trend approaches used here).

Zhang, Q., Harman, C. J., and Kirchner, J. W. (2018), Evaluation of statistical methods for quantifying fractal scaling in water-quality time series with irregular sampling, Hydrol. Earth Syst. Sci., 22, 1175-1192, https://doi.org/10.5194/hess-22-1175-2018.

Specific comments:

6. On Figure 2: a) This is a well-designed figure. b) Consider adding vertical reference lines in the right panel to indicate 1-day, 1-week, and 1-month intervals. c) Add additional reference lines to separate groundwater from stream water – refer to your treatment in Figure 5. d) Consider using color to distinguish between median and mean. e) Comment in the text on the apparent outlier in the site GdQ_198 distribution. f) Do the numbers in bracket represent the number of samples for one constituent or all constituents? Clarify. g) Two of the sites have only one sample each. Justify why those sites should not be removed. In my opinion, those sites which only several samples should also be excluded unless their use can be justified.

7. Line 248: I would suggest using median for the missing value replacement.

8. Line 252: Provide references to justify the use of half detection limit for censored values. It is a typical practice but it has been pointed out that such treatment may cause issues to analysis – refer to the references below. This could be a problem for NO2 and PO4, since the two species have significant proportions of censored values (Table S3).

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

Helsel, D. R. (2005). More than obvious: better methods for interpreting nondetect data. https://pubs.acs.org/doi/pdf/10.1021/es053368a.

9. Line 262: How was the threshold of ‘50 samples’ chosen? It is still a small size.

10. Line 386 (Eq. 2): Check whether you want to use two equal signs in this equation.

11. Line 421: The effect of autocorrelation on trend analysis is not only relevant to short-memory processes (e.g., AR(1) in Yue et al., 2002), but also long-memory processes (e.g., ARFIMA).

Cohn, T. A., and H. F. Lins (2005), Nature’s style: Naturally trendy, Geophys. Res. Lett., 32, L23402, doi:10.1029/2005GL024476.

Zhang, Q., Harman, C. J., and Kirchner, J. W. (2018), Evaluation of statistical methods for quantifying fractal scaling in water-quality time series with irregular sampling, Hydrol. Earth Syst. Sci., 22, 1175-1192, https://doi.org/10.5194/hess-22-1175-2018.

12. Line 456: I think it should be 42% (per Table 2).

13. Line 459: In addition to temperature, PO4 is also negatively correlated with PC 1.

14. Line 463: This should be 18% (per Table 2).

15. Line 537: Check the label for $n < 3$ in Figure 5, which should not be identical to $n < 13$.

16. Line 675: This conclusion should be supported by some references.

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