Review of hess-2019-342: "A predictive model for spatio-temporal variability in stream water quality"

by Guo et al.

Content

This paper introduces a Bayesian hierarchical model for spatio-temporal prediction of water quality variables in Australia. After model construction and validation, the results are discussed in terms of influences on prediction accuracy and regarding the influence of a long drought period on average suspended sediment concentrations. The paper concludes with recommendations regarding model improvement.

General comments

Generally, the paper is well written and the methods and results are interesting. However, I have some major concerns regarding (i) the statements drawn from the results, (ii) influences on the simulation accuracy and (iii) the focus of the study. These major points need to be clarified before publication.

Focus of the study

The study is introduced as a new model for water quality prediction. It is mentioned that the construction of the site-specific model was already published in two preceding papers (Lintern et al., 2018b; Guo et al., 2019). It is not really clear which additional information this paper provides. In the discussion section, there is a long chapter about the influence of a long-term drought to TSS concentrations, which was found as a by-product (?) of the study. The papers ends with conclusions suggesting higher-frequency sampling data, which was not analysed in this study at all. Thus, the study lacks a clear focus and coherent conclusions.

The influence of LOR on simulation accuracy.

- First of all: What is LOR (Limit of Reporting)? Is it a limit of detection (LOD) or a limit of quantification (LOQ) or something different? Which value was used for the calculation of Nash-Sutcliff (Neff) efficiency if the measurement was below LOR? Zero? Half the LOR? Please clarify.

- For model construction, the values below LOR were excluded due to statistical reasons and due to the fact that these low concentrations were of less interest. Thus, why were the values below LOR included in model validation at all? Please clarify.

- Later on it is analysed that the fraction of LOR on total measurement values influences model performance, especially the P fractions and TSS. The discussed reasons are mainly methodical/statistical. I think, the effect of LOR on model performance might also be a secondary effect: the parameters with a high proportion of LOR are mainly those with the highest natural concentration variability, since their concentration peaks are event-driven. Thus, monthly grab samples might capture peaks or not. Since some of the catchments are as small as a few km², even the specific time of a day might influence the sampled concentration to a large extent. Thus, the probability of sampling low between-event concentrations is higher for P and TSS than for e.g. Nitrate. Therefore, the low model performance might rather be an effect of the overall lower information content of the samples, which results in models which are based on a lower information content. What do you think?

Influences of drought on TSS

During the modelling process, the authors note, that a long-term drought influenced TSS concentration, which is a really interesting observation. However, I do not understand why a model is required for this analysis. Wouldn't simple statistics (such a t-test or Mann-Whitney-U-Test) have done the same job? I don't see that this is a special result of this model application.

Meaning of factors.

Since the model is a (multidimensional) statistical model, the explaining variables (factors) not necessarily contain process-based meaning for the target water quality parameters. For example, the water temperature is an explaining variable for temporal variability of TSS (Table S6), which is not really clear to me. In L.15-17 it is stated that the paper addresses the key controls (factors) explaining water quality variability, but an in-depth analysis and discussion is missing in the text. I would encourage the authors to even discuss the factors in more detail or to change the focus of the paper.

Specific comments

L. 1-2: The title "A predictive model for spatio-temporal variability in stream water quality "suggests a generic model for different sites and different water quality parameters. However, the described model is very site-specific. Thus, I would suggest to change the title to a more site-specific one, probably including the region or similar, including the applied method.

L. 71: Change "…quality can not…" to "…quality not…"

L. 76: Change "…model built…" to "… model was built…"

L.76-78. It is stated that the model was constructed and published in two previous papers. Please elaborate on the additional information this paper provides.

L. 79: It is stated, that this study aims at bridging the gap between fully distributed and statistical models. Well, what is this model if not a statistical model? Probably, it was meant to bridge the gap between fully/semi-distributed and lumped models.

L. 154-156. During the Box-Cox transformation of the data, the high sampling values lose their significance, especially for goodness-of-fit calculations. This effect can be seen after back-transformation (figure S13), which results in low Neff values. Thus, how is the statement "poor water quality conditions…were our primary concerns…" compatible to the fact that the data was transformed?

L. 159. Insert a blank between "as each"

L. 186. "… via a Spearman correlation analysis" (note the typo "analyses"). Please add the correlation coefficients and the p-values in the supplement.

L. 246. "…in Sect. 4.2." Isn't it section 4.1?

L. 265. Fix "… is also show…"

L. 414. Fix "For examples, …"

L. 418 Fix "adjscent"

L. 449-451. In the beginning, this paper aims at introducing a model. In this lines, the reader has the impression, that the main aim of this paper is the analysis of drought on TSS concentrations. Please think about the focus of the paper.

L. 466-469: "1) collection … in the model". These are not a results/conclusions of this study. Data frequency was not evaluated in this study.

L. 469-470. "These improvements will be very helpful…" How?