Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-342-RC3, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A predictive model for spatio-temporal variability in stream water quality" *by* Danlu Guo et al.

Mark Honti (Referee)

mark.honti@gmail.com

Received and published: 29 August 2019

GENERAL COMMENTS

The study describes a Bayesian statistical model of selected water quality variables in 102 catchments. The model successfully described both the spatial and temporal variability of certain variables, and performed quite well at describing the site-specific means for all variables. Based on the results, the model can serve as a valuable prediction tool in the calibration region (and potentially adapted elsewhere too).

The main issue with the manuscript is that the otherwise valuable work is presented in an unsuitable (and constantly evolving) context. The title appropriately focuses on the main element of the study, the model and emphasised predictions as the primary

field of utilisation. In the Abstract the motivation for the study is summarised as: "To address this [knowledge gap compromising present water quality models], we developed a Bayesian hierarchical statistical model to analyse the spatio-temporal variability in stream water quality across the state of Victoria, Australia." This shifts from predictions to analysis and promises that the model will cover knowledge gaps presumably by revealing so far unknown relations between water quality and its drivers. Interestingly, this objective is not featured in the Introduction. There it reads: "Our approach aims to bridge the gap between fully-distributed water quality models and statistical approaches to provide useful information for catchment managers, especially for large-scale water quality assessments." This alters the context again, now the model is meant to be a "missing link" between very detailed (deterministic) models and simple statistical tools and the raison d'être is to serve catchment managers. These context shifts do not help to assess the values of the study and generate expectations that are not fulfilled later.

Unfortunately, none of the above alternative contexts is completely followed in the Results and Discussion. The Results consist almost exclusively of performance indicators calculated and plotted in transformed scale. The Discussion focuses on the effects of the drought period on model performance and future development directions without mentioning potential major obstacles and pitfalls (gathering more detailed data and developing more detailed models is an idealistic recipe). The manuscript would greatly benefit from following a clearly defined logical structure, objectives and featuring topics that are truly relevant for the work.

Performance indicators should not occupy all the Results section. There is much more to show about the model, especially considering the ideas that show up in the present Introduction and Abstract. The potential topics include:

- Untransformed comparison of measured and modelled time series for selected catchments

-The needs of catchment managers with respect to predictions and how this model fulfils them (If it does so. If not, management should not be emphasised so much).

-Key controls and mechanisms governing water quality. What do we learn from this study compared to Guo et al. 2019 and Lintern et al 2018a, 2018b?

-The grade of intrinsic randomness (and its compatibility with management), predictability of water quality variables.

-Model limitations: implicit assumptions, conditionality on the calibration set and the present layout of calibration units (what would happen if the model was calibrated on merged catchments?)

-Spatial and temporal distributions of validation errors, their relationship with model development alternatives.

SPECIFIC COMMENTS

Lines 15-16: In my opinion it is not the lack of understanding, but the lack of information. The effects of many key controls on water quality are well understood, albeit in an isolated, idealised context. It is clear, for example what certain polluting sources (like a WWTP effluent, a plot of arable land, etc.) do, how different landcover types affect the transport of pollutants along a specified pathway. The problem with the modelling of stream water quality on the (sub)catchment scale is that numerous key factors and controls act together and in practice there is no hope to get relevant information on all/most of them. That's why detailed and dynamic models fail on all components except those that behave quite simply and are not affected by too many factors. The challenge of modelling is to include the relevant factors AND the necessary information about them. So I would rephrase the sentence to mention that despite the long history of research there are too many key controls and very high complexity in both space and time compared to the available information.

Line 16: Even if there would be a lack of understanding (which I doubt, see previous

comment), how would this issue be addressed by a Bayesian statistical model? Statistical models build on covariance instead of causal relations and therefore are rarely suitable for modelling conditions that are different from the calibration dataset in any significant aspect âĂŤ which is the primary objective of most modelling exercises.

Line 20: Please mention how FRP relates to the more commonly known Soluble Reactive Phosphorus (SRP).

Line 21: The abbreviation of "NOx" is not the best choice, as this is a widely known name of the air pollutant group of gaseous nitrogen oxides. Why not "NOi" or something else?

Lines 21-22: Yes, the model described variation, but above an improvement of understanding is promised.

Lines 29-30: How would a statistical model include those mechanisms that govern non-conservative constituents? Such a development would indeed be a major step forward, but it is definitely not trivial.

Line 32: High frequency data often reveal phenomena that are typically not parts of models and therefore model performance further declines.

Line 33: Besides the classical landuse, agricultural activities (ploughing, fertiliser/pesticide application, livestock handling practices, etc.) would need to be known too.

Lines 40-42: Unpredictable variability does not preclude management. Robust measures can address issues without having to predict the full dynamics. It is well known that the elimination of pollution sources and artificial hydrological factors improves water quality. If the statement in lines 40-42 was true, water quality management would not exist yet.

Lines 42-46: This is a bit lengthy description of the high variability in both space and time. Please consider compressing.

Lines 46-51: Briefly, there are allochthonous and autochthonous emissions and both are subject to transport. Please consider compressing.

Lines 55-59: This listing is somewhat odd. Emission dynamics are completely missing, others are a bit over-detailed and supported with arbitrary references (is the importance of temperature only known since Robert and Mulholland, 2007?).

Lines 60-62: This sentence contradicts the abstract statement (lines 15-16). Water quality modeling faces high epistemic uncertainty, unpredictable variability stems rather from an information gap than the lack of understanding. And what do you mean here by “larger scales”? And please include why effective policy and mitigation need information on variability.

Lines 66-69: It would be worth to mention that most statistical models have weak explanatory and predictive power and therefore it is difficult to use them for designing management interventions.

Lines 71-72: Please check and fix this sentence, by e.g. deleting “can” or any other way.

Lines 74-80: After mentioning management so many times above, one would expect a brief summary about the requirements of managers against water quality models plus a sentence in the objectives on how the current model would fulfil these.

Line 103: Please fix “Beyesian”.

Line 112: Please delete “however”. Either you describe data processing or not. The present formulation suggest that you don’t want to describe it, but later âĂŤ reluctantly âĂŤ still do so.

Line 132: Please briefly mention the forms and indicators of landuse considered among the drivers, because these are non-trivial.

Line 143: You mean “area-specific streamflow”? Streamflow also has the unit of volume

/ time.

Lines 144-149: How did you convert 2D climatic data to soil moisture? This must have included a complete soil hydrological model, but no hints are given in the main text.

Lines 156-157: Low flow days often mean the periods of concern with regard to water quality. What was the case here?

Lines 162-166: This means that you conditioned the transformation on the dataset. Since the predictive nature of the model is emphasised, please explain the procedure of including new catchments. What to do when the new data suggest a different transformation parameter?

Lines 172-174: A random forest approach could have been an alternative for the selection process.

Lines 179-183: Aren’t these results? Since management is emphasised in the introduction, how would you reflect on the final set of key factors? Climate is close to impossible to manipulate, temperature, soil moisture and streamflow are difficult. Why no direct human factors other than landuse?

Lines 194-196: What is the rationale behind the half-normal prior? What is the advantage compared to an exponential? The half-normal suggests that relatively small standard deviations are equally likely, while the exponential prioritises as small std. deviation as possible. Please justify your choice.

Lines 212-214: This is a rather extreme test, why do you expect the model to describe the below-LOR data, after excluding all of them from the calibration dataset. Would a good fit mean that below-LOR data follow the same rules as above-LOD data do? Line 217: The verb “suggested” sounds weird to me here.

Lines 238-240: FRP is a subset of TP. TP has complicated relations to TSS. The FRP-TP relationship is governed by several (fast) biochemical processes simultaneously. Consequently, it is no surprise that FRP is hard to model without considering all these intricate interactions. By the way, a negative NSE suggests that the model entirely failed to capture any of the real dynamics (negative NSE means that a constant model at the mean would perform better).

Figures 2-3: It would be great to see some visualisation beyond 1:1 plots in transformed space (of unknown transformation parameters unless one digs them up from elsewhere).

Lines 268-269: This sentence is not necessary, the section title tells the same.

Lines 269-270, 273: Please delete the “Note that . . . in Sect. .3.1.” sentence and add “We exclude the FRP model from the analysis due to its poor performance (section 3.1).” into Line 273 after “monitoring sites.”.

Tables 1-2: These tables are all about calibration indicators, and not the subject of the model. These could be moved to the SI. Why not showing something about the factors? The introduction promised filling some knowledge gaps yet we do not learn about anything except performance indicators (and later the influence of drought on them in table 3).

Line 324: The Results section is over, yet the roles of “key controls”, the proportions of “inherent randomness” both remain untold. The primary value of such a model is its information content, which is embodied in the relationships that turn inputs to outputs using the parameters. Model performance indicators are important too, but in a secondary sense: they help to assess the quality of information that can be obtained from the model. Here the reader learns about the model performance in various cases, yet the lesson can’t be learnt. What governs the different water quality variables? Are there covariations between the variables? Are certain models similar to others? Are errors clustered in certain situations? Which environmental factors influence the variables, how sensitive are they to the most important one? Etc.

Lines 333-334: Would be more positive to start with the opportunities and afterwards

with limitations.

Lines 334-335: Or when the variability is high and explanatory power is weak. Very low FRP values could be much better simulated given that the model knows all the influencing factors and processes.

Line 336: This can also be by chance. TKN and EC are “more conservative” than the others, and have much weaker relations to sediment.

Lines 347-359: It is true that transformation increases the distance between distinct values close to the numerical resolution of data, which violates the linearity assumption. But when you do not transform, linearity is violated by default (as one of the aims of transformation is to reduce nonlinearity). Besides the alternative model structures mentioned, a practical solution is to perturb the data with random small values (small fraction of numerical resolution), which dissolves the discrete bands of the low values without significantly altering the data. This is basically the same as “measurement noise” beyond the resolution of the time-series.

Lines 360-361: Yes, this was obvious from the start. That’s why the “positioning” of the model study is not optimal. The applied methodology tested whether temporal / regional differences could be replicated by a simple statistical model that lacks any mechanistic background. The exposition of knowledge gaps, management-relevant factors, general predictive power for ungauged catchments create expectations that simply cannot be fulfilled by this model. A lot of mechanistic knowledge is available for these water quality variables, no single bit of this knowledge is reflected by the model structure. A more realistic context would have been to investigate the overarching patterns in this region of Victoria, emphasising that the model only considers emissions only implicitly, through landuse, which in turn assumes similar human activities in the same landuse type. The results are completely in line with previous experiences, more conservative and less sediment-related variables are easier to predict than the others. The model can be a valuable predictive tool, but only in the region of calibration and only for those water quality variables, for which have the model performed acceptably.

Lines 364-369: Making the model more detailed can potentially lead to a dead end. Non-linear statistical model structures may perform a bit better, but need more data for a meaningful calibration and still often lack the mechanistic background, and are much more complicated numerically. Adding descriptions of different mechanisms to the model either moves it towards a deterministic direction, which is a wrong way for this spatial and temporal scale because data will anyway appear to be at least partly random due to the lack of information on all relevant drivers, or leads to a stochastic-dynamic model, which is extremely complicated and difficult to calibrate.

Lines 372-373: If this was an issue, why don't we learn about the "real-world" (=non-transformed) model accuracy earlier? The NSE values and the figures are all in transformed space, so it is difficult to judge what these mean for the practice.

Lines 375-377: I don't understand this example. Completely usual floods often bring much more sediments in almost pristine mountain catchments. Why would such an event be an alarm for management?

Lines 377-379: How? This should have been the main topic if the logical line of the Introduction was followed. How strong is the predictive power of the calibrated models considering practical needs? Are they suitable for real forecasting either for the far future or for shorter periods during operative management?

Lines 384-386: The references are "too new" for this statement in its present general form. Commercial solutions for online monitoring with <10 minute resolution is available for turbidity (proxy for TSS), temperature, EC, chlorophyll, dissolved oxygen since at least 20 years. Nutrient sensors are indeed newer, yet they are often not sensitive enough to yield meaningful data in surface waters (unless they are heavily polluted).

Lines 386-388: How would you apply remote sensing in stream networks? Except for larger rivers (and of course, lakes and reservoirs), these water surfaces are difficult to

analyse because the number of "clean" pixels without any terrestrial or littoral influence is very low or even zero.

Lines 390-391: There may be better (older, original) references for this. This is known since at least 30 years.

Lines 391-397: Please remove, this is too case-specific.

Lines 398-399: This is the exact reason why models fail despite the rather solid understanding of mechanisms (and this is a data or information gap and not a knowledge gap). Relevant, representative, and accurate data on such activities is close to impossible to obtain, even for smaller regions or shorter periods. Therefore, the temporal and spatial variability of these contribute to apparent "inherent randomness" and undescribed variance (the difference of NSE from 1) and weaken the predictive power of models. At the moment the solution to this issue remains an open question even for the past/present, not to mention the potentially changing practices of the future.

Lines 422-423: Direct livestock input may increase concentrations during drought.

Lines 438-443: As the results of this study showed, this would be a hard job without implementing at least a few mechanistic features in the model. However, more features would require more data, potentially beyond the scope of the presented dataset.

Lines 455-457: This is a crucially important sentence. I would add explicitly that the model is not only bound to the period, but also to the region for which calibration took place.

Supplementary material: Figures could be structured better graphically. When 4x4 panel units are to be seen, please structure the figure so that the units get obvious. Please indicate the contents in the subfigure title. Print Box-Cox or log-sinh transformation parameters on figures or in the caption, because without knowing the strength of transformation it is difficult to judge the quality of fit.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-342, 2019.