Responses to Comments on "A Bayesian approach to understanding the key factors influencing temporal variability in stream water quality: a case study in the Great Barrier Reef catchments" (Referee #1)

Anonymous Referee #1 Received and published on 10 Feb 2021.

Our responses are in blue and proposed manuscript revisions underlined.

General comment

This manuscript presents a Bayesian modeling approach to understanding factors affecting temporal variability in stream water quality. Overall, I think the manuscript is well written and will become a worthwhile contribution to the hydrological community after moderate revisions. Below I provide some comments to the author, which I hope can help improve the manuscript.

The authors acknowledge the referee's positive comment and the recognition of contribution of this study. The constructive comments will help us improve our manuscript after revision. We provide detailed responses to your comments and our proposed manuscript revisions in the subsequent sections.

Specific comment

1. The authors have made it explicit that the current work follows previous study investigating water quality variability in the same region (Liu et al., 2018). There are also other publications from these authors, e.g., Guo et al., 2019, 2020. The discussion section seems not provide much comparison or synthesis of the results from these different but related studies, which appears to be a missed opportunity. I am aware some of these studies focused on temporal patterns and some on temporal patterns. It can potentially become a nice addition to the manuscript and a contribution to the community if the authors can provide some reflection on what different modeling techniques they have used and what new insights on water-quality patterns they have learned from those techniques.

The authors agree with the reviewer and we will incorporate this suggestion. The innovations that this study brings compared to previous studies: 1) Queensland is more event dominated, thus we used event-based water quality data, compared to our previous studies which used monthly water quality data in Victoria; and 2) different modelling methods are used in this study - we used a model averaging approach, rather than a universal modelling framework, which is a more robust approach to understanding the key factors, since the effect of key factors are derived from multiple models.

<u>To address this comment, we will provide more synthesis of the results from this paper, comparing with previous water quality modelling studies. This includes:</u>

1) we will highlight the position of the current study to the broader water quality modelling community in ***Introduction***. The focus of this part is to demonstrate different modelling approaches (e.g. simple regression models and process-based hydrological models) (Bartley et al., 2012; Hirsch et al., 2010; Khan et al., 2020; McCloskey et al., 2021), and how these approaches address the spatial and temporal patterns of water quality (Barrientos et al., 2018; Hrachowitz et al., 2016; Kaman et al., 2016; Varanka et al., 2015).

2) we will provide more discussion that specifically compare this study to our previous papers (Guo et al., 2020; Guo et al., 2019) in ***Sect.4.2 (Predicting temporal variations in water quality)***.

2. The authors have analyzed nine water quality constituents. While I do appreciate the amount of efforts the authors invested in data analysis and modeling, I wonder if it helps everyone stay focused if the authors were more selective on the constituents. Since a key message from this work is on the different drives of particulate and dissolved constituents, it may be sufficient to select two constituents from each category, as opposed to showing data and results for all nine constituents.

The authors appreciate this suggestion. While it might help the reader more focused when only selected constituents were included in this paper, we would like to keep all nine constituents in the revised manuscript. The reasons that support this decision include:

1) There is a large number of constituents that have been monitored in the GBR water quality monitoring program, but we have reduced the number of constituents for those that have similar patterns. For example, we only retained TSS among TSS, total nitrogen (TN) and total phosphorus (TP).

2) Our analyses are on 9 constituents that are of great concern to the coral reef ecosystem (McCloskey et al., 2017), and could provide a useful comprehensive picture on the overall water quality status and its key temporal drivers. We only consider the 'real parameter' that can be directed measured (expect NOx). This helps to understand full sediment and nutrient budgets exported to the GBR lagoon.

To resolve this comment, in the revised manuscript, the following proposed changes will be made:

1) We will highlight the reason why we select these nine constituents in ***Introduction***.

2) We have already been selective on presenting results, focusing on TSS, NOx and FRP i.e. one constituent per category. We will explain in the paper that the results have been simplified and explain our constituent selection rationale in ***Sect. 3.1 (Key drivers of temporal variability in water quality)***. In addition, a number of the graphs are already only done for three constituents (e.g., Figures 5 to 8), and we will simplify other figures and tables (e.g., Figure 9 and Table 4) to reflect our focus.

3. Of the two clusters of sites (Figure 2), Cluster 1 sites are quite concentrated, whereas Cluster 2 sites are much more scattered. Also, there seems to be more sites in cluster 2 than cluster 1. I noted that the Bayesian modeling framework was applied to the two clusters independently, I wonder if any of these two aspects (geographical proximity and number of sites) could potentially affect your models and comparison of results between the two clusters. In addition, have you considered developing a single Bayesian model on all sites with the cluster assignment has an explanatory variable?

Thank you for this comment. The clustering results are based on our previous multivariate analysis on the spatial pattern of water quality in the same study area (Liu et al., 2018). We found that distinctive features of the two clusters and their geographic/hydroclimatic differences are responsible for the separation of two clusters of sites. For instance, small wet areas (Cluster 1) near the coast where topography (orography) plays an important role in rainfall generation. Such geographic features also lead to more dispersed sites in the drier area (Cluster 2).

Furthermore, there are good conceptual reasons for keeping the clusters separate. Based on Liu et al. (2018), results from clustering analyses on spatial patterns of water quality and catchment characteristics were highly correlated, and the two clusters had quite different key explanatory variables. If we put all the sites into the same analysis and just included cluster identity as a random intercept (or even random slope for each explanatory variable) it would skew the choice of explanatory variables for both clusters away from the set achieved in the analysis as it stands. We would end up with same key factors identified for two different clusters, which provides limited information on specific management focuses on two contrasting sets of catchments.

To address this comment, in the revised manuscript, we propose to:

1) provide more details on our previous study on clustering of these catchments in ***Sect. 2.1 (Study area)***, and that differences in geographic/hydroclimatic are key factors that distinguished the two clusters of sites. Thus, there are strong practical merits in handling the clusters separately based on the clear contrast between them.

2) improve our description in ***Sect. 2.3 (Modelling: driver identification and water quality prediction using multi-model inference)***, to further clarify the reasons behind applying Bayesian model averaging on two different cluster separately.

4. Line 35: In addition to sources, mobilization, and delivery, “transformation” should be included.

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

5. Section 2.2.3: The authors have quantified the correlation between explanatory variables (Figure B1). Have you considered excluding some variables based on the correlations? If any two variables are highly correlated, it may be wise to keep just one of them in the models.

We have examined the correlation among all explanatory variables, and there are several pairs of variables that are highly correlated (e.g., pre-event NDVI and event NDVI with Spearman's $\rho = 0.97$). However, it does not necessarily mean they will have similar posterior inclusion probability from BMA (e.g., 1.00 and 0.34 for pre-event NDVI and event NDVI, respectively, for DON in Cluster 2). The BMA can handle the collinearity with shrinking the posterior distribution of the correlated variable to near zero (Posch et al., 2020). This shrinkage effect leads to lower posterior probability of the more complex model (i.e., the model that includes correlated variables), because each extra parameter dilutes the prior density on the pre-existing parameters. Thus, models that include more predictors will have a lower prior probability. Models with additional predictors will be favored only to the extent that their benefit in higher likelihood outweighs their cost in lower prior; however, including correlated variables does not increase the model predictive capacity (Daoud, 2017; Hinne et al., 2020; Kruschke, 2014).

Furthermore, Freckleton (2011) highlighted that when applying model averaging approach, it is not safe to simply exclude correlated variables without due consideration of their likely independent effects. In our case, the high correlation among predictors mainly comes from time lag effects between predictors (e.g., pre-event, event and post-event). The relative importance of these predictors provides strong management indication for future water quality management strategies. Therefore, we would like to keep them all in this analysis.

To resolve this comment, we propose that:

1) highlight that some of the variables are proxies for the same process, and therefore they are closely related in ***Sect. 2.2.3 (Explanatory variables)***. We will pay attention to the collinearity issue in the analysis of the results.

2) add more clarification in ***Sect. 3.1 (Key drivers of temporal variability in water quality),*** that strong correlation between predictors does not necessary mean that the posterior inclusion probability of these factors is similar. In addition, we will provide more discussion on how BMA address the collinearity issue in in ***Sect. 3.1.***

6. BMA model coefficients plots (Figure 5 and other related figures in the SM): I found it difficult to compare the patterns across clusters or among constituents because the variables are not displayed in the same order in these panels.

Thank you for this comment. We will incorporate this suggestion in the revised manuscript (e.g., reorder the predictors of Figure 5 and other plots in SM to make sure they follow the same order).

7. Predictive model performance (Section 3.2 and Table 4): The NSE values are not high, some are very low. This seems to limit the utility of the proposed Bayesian approach, which the authors should discuss and defend against.

We agree with the referee that the NSE values are not high, but based on the recommended performance measures from Moriasi et al. (2015), most of the model performance is satisfactory (Table 1), especially for the Cluster 2 models. Generally, low NSE is acceptable for modelling nutrients and sediment compared to hydrology. It is also worth noting that most of water quality models evaluated in Moriasi et al. (2015) are physically-based models (e.g. SWAT, HSPF, WARME), and focusing on individual catchments. However, we used a statistical modelling approached to predict multiple catchments and to identify key factors simultaneously. We agree that the model performance for DOP in Cluster 1 is very poor, and we have provided detailed discussion of this in ***Lines 428 to 438*** in ***Sect. 4.2 (Predicting temporal variations in water quality)***. Therefore, we did not rely on any results for DOP in Cluster 1 when analyzing the results.

Table 1 Performance statistics for nine constituents for the modelled and observed temporal variability, according to Moriasi et al. (2015).

Constituent	Cluster one	Cluster two
TSS	Indicative	Satisfactory
PN	Satisfactory	Satisfactory
NO_X	Satisfactory	Good
NH_4	Satisfactory	Indicative
DON	Satisfactory	Satisfactory
FRP	Satisfactory	Satisfactory
DOP	Indicative	Good
PP	Satisfactory	Satisfactory
EC	Satisfactory	Satisfactory

To address this comment, we propose to:

1) provide additional assessment of model performance based on the recommended performance measures from Moriasi et al. (2015) in ***Sect. 4.2 (Predicting temporal variations in water quality).*** We will demonstrate that our predictive ability is comparable to other water quality models.

2) provide more discussion in ***Sect. 4.2*** that we are not inferring any conclusions from the modelling results for DOP in Cluster 1, due to the poor performance.

8. Line 415: Again, the effect is not only on transportation but also on transformation. Specifically, temperature is expected to affect the intensity of biological processes, e.g., denitrification.

Thank you for this comment. We will incorporate this suggestion in the revised manuscript.

Reference

Barrientos, Guillermo, & Iroumé, Andrés. (2018). The effects of topography and forest management on water storage in catchments in south-central Chile. *Hydrological Processes, 32*(21), 3225-3240.

Bartley, Rebecca, Speirs, William J, Ellis, Tim W, & Waters, David K. (2012). A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. *Marine pollution bulletin, 65*(4-9), 101-116.

Daoud, Jamal I. (2017). *Multicollinearity and regression analysis.* Paper presented at the Journal of Physics: Conference Series.

Freckleton, Robert P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology, 65*(1), 91-101.

Guo, Danlu, Lintern, Anna, Webb, J Angus, Ryu, Dongryeol, Bende-Michl, Ulrike, Liu, Shuci, & Western, Andrew William. (2020). A data-based predictive model for spatiotemporal variability in stream water quality. *Hydrology and Earth System Sciences, 24*(2), 827-847.

Guo, Danlu, Lintern, Anna, Webb, J Angus, Ryu, Dongryeol, Liu, Shuci, Bende-Michl, Ulrike, Leahy, Paul, Wilson, Paul, & Western, AW. (2019). Key Factors Affecting Temporal Variability in Stream Water Quality. *Water Resources Research, 55*(1), 112-129.

Hinne, Max, Gronau, Quentin F., van den Bergh, Don, & Wagenmakers, Eric-Jan. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science, 3*(2), 200-215. doi: 10.1177/2515245919898657

Hirsch, Robert M, Moyer, Douglas L, & Archfield, Stacey A. (2010). Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs 1. *JAWRA Journal of the American Water Resources Association, 46*(5), 857-880.

Hrachowitz, Markus, Benettin, Paolo, Van Breukelen, Boris M, Fovet, Ophelie, Howden, Nicholas JK, Ruiz, Laurent, Van Der Velde, Ype, & Wade, Andrew J. (2016). Transit times—the link between hydrology and water quality at the catchment scale. *Wiley Interdisciplinary Reviews: Water, 3*(5), 629-657.

Kaman, Harun, Cetin, Mahmut, & Kırda, Cevat. (2016). SPATIAL AND TEMPORAL EVALUATION OF DEPTH AND SALINITY OF THE GROUNDWATER IN A LARGE IRRIGATED AREA IN SOUTHERN TURKEY. *Pakistan Journal of Agricultural Sciences, 53*(2).

Khan, Urooj, Cook, Freeman J, Laugesen, Richard, Hasan, Mohammad M, Plastow, Kevin, Amirthanathan, Gnanathikkam E, Bari, Mohammed A, & Tuteja, Narendra K. (2020). Development of catchment water quality models within a realtime status and forecast system for the Great Barrier Reef. *Environmental Modelling & Software, 132*, 104790.

Kruschke, John. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*: Academic Press.

Liu, S, Ryu, D, Webb, JA, Lintern, A, Waters, D, Guo, Danlu, & Western, AW. (2018). Characterisation of spatial variability in water quality in the Great Barrier Reef catchments using multivariate statistical analysis. *Marine pollution bulletin, 137*, 137-151. doi: https://doi.org/10.1016/j.marpolbul.2018.10.019

McCloskey, G, Waters, D, Baheerathan, R, Darr, S, Dougall, C, Ellis, R, Fentie, B, & Hateley, L. (2017). Modelling reductions of pollutant loads due to improved management practices in the great barrier reef catchments: Updated methodology and results-technical report for reef report card 2015. *Queensland Department of Natural Resources and Mines, Brisbane, Queensland*.

McCloskey, GL, Baheerathan, R, Dougall, C, Ellis, R, Bennett, FR, Waters, D, Darr, S, Fentie, B, Hateley, LR, & Askildsen, M. (2021). Modelled estimates of fine sediment and particulate nutrients delivered from the Great Barrier Reef catchments. *Marine pollution bulletin, 165*, 112163.

Moriasi, Daniel N, Gitau, Margaret W, Pai, Naresh, & Daggupati, Prasad. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE, 58*(6), 1763-1785.

Posch, Konstantin, Arbeiter, Maximilian, & Pilz, Juergen. (2020). A novel Bayesian approach for variable selection in linear regression models. *Computational Statistics & Data Analysis, 144*, 106881.

Varanka, Sanna, Hjort, Jan, & Luoto, Miska. (2015). Geomorphological factors predict water quality in boreal rivers. *Earth Surface Processes and Landforms, 40*(15), 1989-1999.