Responses to comments on: "Daily hypoxia forecasting and uncertainty assessment via Bayesian mechanistic model for the Northern Gulf of Mexico" (Referee #3)

Our responses are in blue.

General comments

This is a well written manuscript on an extremely important topic of predicting hypoxia in a water body of global significance. The methods and findings from this study are also applicable to similar water bodies across the world. The novel approach of using data-driven modelling (particularly, Bayesian statistical modelling) makes this a suitable paper for HESS. The results and discussion are generally well written (save one comment about being confused by Table 2).

We are grateful for the referee's positive comment and the recognition of contribution of this study.

However, I do think that it would be good to see more details in the methodology about the approach taken in this study. This is coming from the perspective of someone who thinks this work is very relevant to the work of watershed planners and managers – who may want more detail on how they could apply similar work themselves.

I understand the reasoning for not detailing the Bayesian Mechanistic Model in this paper (because it was previously published) - however, I do think that more details on the model formulation are required in Section 2.1 so that this can be a stand-alone paper in its own right. Particularly - what are the key parameters and what are the model equations? Similarly in Section 2.2, readers need to refer to the DMO 20 publication, and to Matli et al. 2018. I do think that these details should be brought across to this paper too.

Thank you for the comments and suggestions. In this manuscript, we focus on the new forecasting approach, but not the process-based model, which was described by DMO20. To address the reviewer's concern, we will add the main DMO20 model equations and a summary of the Bayesian posterior parameter estimates to the Supplementary material. We will also provide more explanation of Matli et. al., (2018) in the methods (Section 2.2).

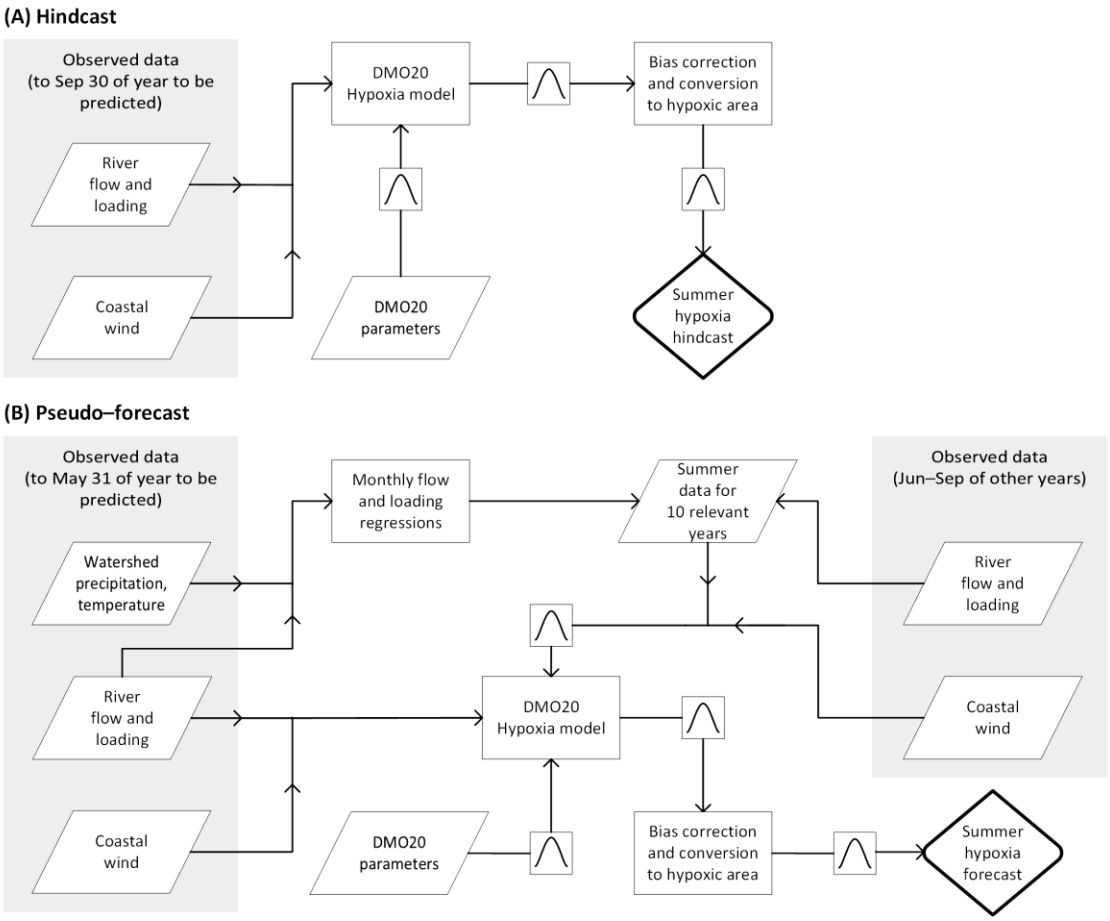
I also think more emphasis in the methodology needs to be placed on the forecasting method and the regression modelling of June-September Discharge and Loading. I didn't realise until quite later in the paper that Bayesian methods were used in determining these data. This is quite important and could be applied not only in the context presented in the paper, but to other sites. In particular the details I would like to see are: what software/platform was used for the modelling (is this available on github etc?), how did you check convergence of chains?, how many iterations?, was there a burn-in period?, how many models did you produce in the exhaustive search (just to place the scale of the work in context), what are the prior distributions?, how were the assumptions checked?

We appreciate the comment and would like to point out that we use conventional linear regression (not Bayesian) to predict June–September discharge and loading. We preferred conventional over Bayesian methods due to the lack of strong prior information and higher

computational efficiency in the exhaustive search application. For situations without prior information, Bayesian and frequentist predictions are generally equivalent. We describe the process of constructing the regressions in Section 2.4 and results of the regressions in Section 3.1 and the observed vs predicted plots (Figures S3.2–S3.5). We did use the "Bayesian Information Criterion" for variable selection, but this is really more of a frequentist method, despite what the name might imply. We report the relevant R packages and methods in the Methods, and we will add a general reference to R (the software platform used in this study).

I would also like to see an assessment or further discussion of how the uncertainty in the forecasts are propogated through the Bayesian Mechanistic Model. How do you account for this? How does the final uncertainty change? More details on how this assessment is conducted would be good to see in the methodology.

Thank you for the comment. In order to improve the description of uncertainty propagation, we will update Figure 1 with symbols that indicate linkages where uncertainties are propagated through the forecasting procedure. In addition, we will add boxes for the regression conversions to Figure 1. Please see the draft of the updated Figure 1 below this response. We think Figure 4 and the associated discussion provides a useful assessment of the various uncertainties accounted for in the forecasting approach.



Specific Comments:

Lines 60-63: I am a little confused as to what the difference here is between hindcasting and forecasting? Would these need different modelling/simulation strategies? Is the forecasting process the same as the hindcasting process - just using input data representing future conditions? Then how would you validate the success of the forecasting? Perhaps a brief clarification of these points would be useful here.

Thank you for the comment. Hindcasts represent model predictions of past conditions assuming all input data are known throughout the prediction period (e.g., flows, loads, and winds are known throughout the summer). Forecasting, on the other hand, requires us to make predictions and/or assumptions regarding some model inputs (e.g., flows, loads, and winds after the forecast release date of 1 June). We explain the forecasting process in the Methods Section 2.3. To improve the differentiation between forecasting and hindcasting, we will expand Figure 1 to include a flowchart for hindcasting (see draft above). The most compelling validation of the pseudo-forecasts is comparison to the geostatistical "observations" (which are based on monitoring cruise sampling data). We also compare pseudo-forecasts to hindcasts, since hindcasts represent the best model estimate of historical hypoxic conditions.

Line 101: 'geostatistical estimates from Matli et al' - please elaborate on this a bit more: what are the geostatistical estimates of? What does Matli et al. 2018 provide?

Please see our response to the next comment.

Line 101-102: Could you please provide a few more details here too? I assume that this is referring to the estimates of BWDO and HA, but it is not 100% clear. Why is it that the monitoring cruise data have lower uncertainty? Also - how often do these monitoring cruises happen?

We will expand the description of these geostatistical estimates. In general, there are 149 cruises across our study period: 34 in June, 63 in July, 35 in August, and 17 in September. Geostatistical estimates are based on the cruise observations, such that there is greater uncertainty if/when trying to interpolate temporally (between cruises).

Line 321-324: The authors state that the projections of summer riverine inputs improve hypoxia forecasting skill - and refer to Table 2, however even after the explanation in Section 3.2, I am still confused about this table. I would have thought that if the projectings are improving forecasting skill, we should see higher R2 in the Forecasted vs Observed columns compared to the Forecasted vs Hindcasted columns. Or perhaps I am competely misunderstanding the table.

Thank you for this comment. Table 2 presents a comparison of using varying riverine and meteorological data inputs, as described earlier in the manuscript (Section 3.3). The purpose of Table 2 is to compare the four different forecasting input cases (focusing on a comparison across table rows, not columns). We will clarify this in the text by referencing Case 4 (best case) at Line 322. Forecast performance generally correlates more strongly with hindcasts than observations (because both forecasts and hindcasts are based on the same mechanistic hypoxia model). We also hope that our revised Figure 1 will help resolve some of this confusion.

Technical Comments:

- The authors have used a large number of acronyms throughout the manuscript (e.g., HA, BDOM, NGoM, MAR). This is perhaps a subjective comment, but I highly suggest that these acronyms are avoided and terms written out in full to make the manuscript more readable.

In the spirit of this comment, we will remove the NGoM and MAR acronyms.

- Fig S1: I suggest putting this figure in the main manuscript - not all readers are familiar with the Gulf of Mexico

Since there have been many publications on the northern Gulf hypoxic zone, we do not think it is critical to include this map in the main manuscript. However, if the editor agrees this is important, we will add it.

- figure 1 was a very useful figure for me and helped me understand the process - but for those who are not familiar with the different text box shapes - could you please provide a legend either in the figure or in the caption? also perhaps to make it easier for the reader, it might be nice to have the section heading numbers in the figure.

We appreciate the comment and will modify Figure 1 (see draft above). We have simplified the shapes, and we will also update the figure legend and/or caption.

References

Matli, V. R. R., Fang, S., Guinness, J., Rabalais, N. N., Craig, J. K. and Obenour, D. R.: Space-Time Geostatistical Assessment of Hypoxia in the Northern Gulf of Mexico, Environ. Sci. Technol., 52(21), 12484–12493, doi:10.1021/acs.est.8b03474, 2018.