Responses to comments on: "Daily hypoxia forecasting and uncertainty assessment via Bayesian mechanistic model for the Northern Gulf of Mexico" (Referee #2)

Our responses are in blue.

Comments

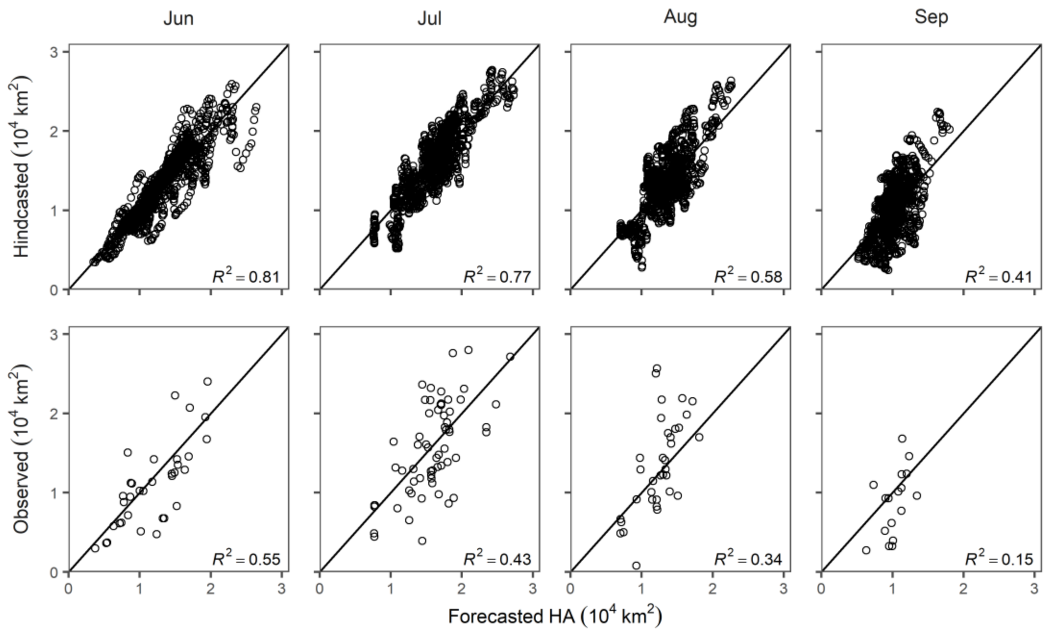
This manuscript builds upon previous statistical models for hypoxia in the Gulf of Mexico to make a temporally resolved forecast. This work is potentially valuable for management of the ecosystem's fisheries and offers some important insights regarding the contributions of different sources of uncertainty to this type of forecast. The manuscript is well written and will be of interest to readers of HESS. I have one substantial concern and a number of comments for the authors to consider.

We appreciate the positive feedback and acknowledgement of potential utility to management.

This modeling approach is impressive as it handles the different sources of uncertainty without excessive parameterization or an actual process-based model. To do accomplish this, the model relies on use of historical data for the trajectory of summer meteorology, discharge and loading. The authors select the most relevant years' summer records based on similarity of spring forcings. This selection process at line 115 is explained clearly, but more justification would be helpful as this is potentially influential for the forecast. What are the consequences if this selection step were omitted? What happens to predictive performance if 10 less relevant years are used? Does that lead to degradation of performance?

Thank you for the positive comment on forecasting approach. A tighter selection (with fewer relevant years) may produce more accurate forecasting results, but may fail to capture the true variance (stochasticity) in the hydrometeorology. Thus, 10 years was selected as a reasonable balance between accuracy of prediction and accuracy of uncertainty characterization. To explore this more, we ran the forecasting procedure selecting only 5 relevant years, and the predictive accuracy was slightly improved (compare figure below to Figure 3 of manuscript). However, we estimate the confidence interval for the population variance to increase by a factor of 1.5 when using 5 years instead of 10 years. We will add a discussion of this analysis to the Discussion section of the manuscript.

If we omit the selection process entirely, performance degrades substantially. This is actually explored in Section 3.3. In particular, compare Case 4 to Case 2 in Table 2.



It appears that this temporally resolved forecast would, in fact, be static once the spring hydrology data are used to identify the most similar years to use for summer forcings. Could this somehow be informed by additional data from after May, perhaps using some information from previous years?

In general, we agree that it could be interesting to update the forecast over the summer, as additional data become available (after May). While outside the scope of this study, we will note this as potential future research in the Discussion. However, we are unsure what the reviewer means by “using some information from previous years”, as we already do this. In addition, we will update the title of the manuscript to “Temporally resolved coastal hypoxia forecasting and uncertainty assessment via Bayesian mechanistic modeling” clarify the main point of this work.

Figure 2. Are the hindcasted data for all of the years, or only those that were matched as most relevant and used for the model?

This figure shows results for all years. Hindcasted data represents the predictive output of DMO20 using the actual hydrometeorology (assuming it is known throughout the summer). Thus, “relevant years” aren’t relevant for hindcasts. The pseudo-forecasts are based on relevant years (and the relevant years change for each year being forecasted). We will edit the figure caption to clarify. Note that we will also edit Figure 1 to include both the hindcast and pseudo-forecast procedure, which should help to further clarify the methodology for the reader. The draft of the revised Figure 1 is shown below.

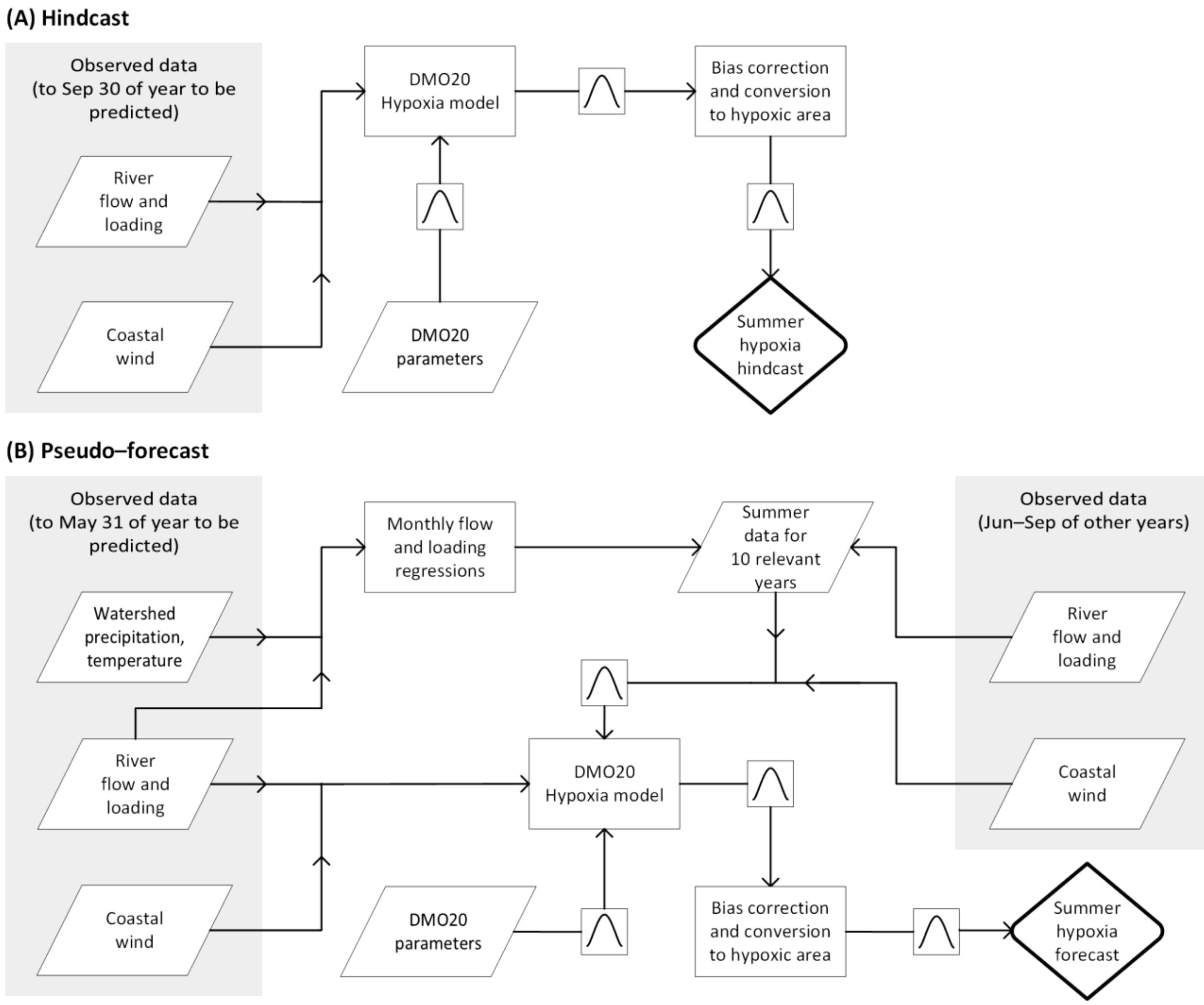


Figure 4. The predictive intervals in the shaded region do not appear to increase over the course of the season, which is unexpected based on Figure 3. The authors provide parts of the explanation for this in lines 240-245, but I suggest adding an additional sentence that puts together the 1) disparity in uncertainty between inputs and fitted parameters and 2) change in contribution of inputs over the season.

Thank you for the comment, and we agree that the predictive intervals in Figure 4 may seem unexpected. This is largely because predictive intervals tend to increase with hypoxic area size, and the hypoxic area size tends to decline after July. For this reason, we included Figure 5 and corresponding text at lines 258-268. As shown in Figure 5A, normalized predictive intervals (normalized by hypoxic area) do increase over the summer season (though they are still fairly high in mid June due to strong and highly variable wind speeds, as shown in Figure 5B). We will edit the text here to further clarify this point.

Table 2 and Line 326. The caution about increasing uncertainties over the season is appropriate, but is not adequately captured in the figures or tables in the main text or supplement. Instead, the decrease in R^2 is presented. Although the uncertainties are referenced in the text, it would be helpful to have a table or visual that shows the uncertainty for the pseudo-forecast by month (by that I mean IQR in units of HA or BWDO)

Thank you for the comment, however, we do not think the additional table is necessary as we discuss the increase of uncertainty in later August–September in the Lines 260-262. Besides, the uncertainty increase in later months can be seen in Figure 5. As noted in our previous response, we will further clarify this in the text.

Technical Comments

Table 2. What does the bold styling indicate? It appears to be the highest R2 of the four cases, but this could be explained in the caption

Thank you for the comment, the bold styling indeed indicates the highest R^2 of the four cases (by month) and we will add clarification in the table caption.

Line 91 – The rationale for this correction is clear and sound. Presumably there is some uncertainty associated with both the predictor and fitted parameters of this regression. Were those carried forward into the forecast? I would expect that uncertainty to be impactful.

We agree with the comment and most likely the uncertainty for early June forecasted values will increase. We will add the uncertainty associated with this regression to the model and update the manuscript. However, we do not expect it to considerably increase total forecasting uncertainty.

Line 135 – Did this occur frequently? Assigning a threshold for performance is appropriate, but consider adding some justification for why this particular threshold was selected.

Thank you for the comment. All the final regression models and their performance are presented in Table 1. As shown in Table 1, there is a large drop in performance between July and August (as noted in Lines 174-175), so results would be robust to moderate increases in this threshold. Also, we performed leave-one year-out cross validation (see response to Reviewer #1), which suggested even poorer performance of August–September regressions in projecting both flow and loading. We will update Table 1 with leave-one-out cross validation results and add discussion supporting our decision.