Responses to comments by Referee #3

Comment on bg-2022-4
Anonymous Referee #3

The manuscript employs a novel approach in applying numerical and statistical modeling techniques to more accurately forecast hypoxia area on the Louisiana-Texas shelf in the Gulf of Mexico. After selecting a set of predictors that are well correlated with hypoxic area in the Gulf, a long-term ROMS numerical simulation of this study area (2007-2020) is used to train an ensemble of statistical models using both generalized linear and generalized additive modeling techniques. The most promising techniques are then applied to global model outputs and USGS forcings to develop an accurate forecast over a later time period (2019-2020).

Overall, the manuscript describes a highly applicable and useful approach to rapidly forecast hypoxic conditions using a statistical ensemble. This approach appears to offer multiple benefits to past forecasts and would serve as a helpful template for other coastal areas as well. The paper utilizes a limited number of explanatory variables to achieve a good fit, and I think that the predictors they use are appropriate and highly applicable to hypoxic area estimates. I've tried to include many notes to summarize these points, but this is not an exhaustive list.

Major comments

General:
There is a fair amount of general awkward phrasing and minor grammatical and spelling errors, but I don't find that they hinder my own understanding of the content.

Introduction:
I think that this section could be broken up into three sections as opposed to the 2 paragraphs it has now. Currently, only one sentence discusses the ecological/societal consequences of hypoxia in this region, and the authors immediately begin discussing the predictive capabilities of previous forecasting efforts. In my opinion, there could be more motivation in the first paragraph that illustrates why hypoxia forecasts are important and useful, and the benefits that environmental managers and others could gain from an accurate forecast. Otherwise, this reads a bit more like an interesting scientific modeling exercise done for its own sake. The second paragraph could then focus on past efforts to create a forecasting system, while the final paragraph could talk about some of the shortcomings that this model ensemble will address.
Authors' Response: We agree with Reviewer#3 and will rearrange the Introduction following the comment.

Methods:
I have some minor questions about the equations described for the hydrodynamic-related predictors section, but I don't think that they are likely to alter the conclusions of the paper in a meaningful way.

Discussion:
Would suggest renaming this section as "results" since a discussion section is typically what is described in the conclusions section here.

Authors' Response: We will move the application of HyCOM dataset to the Result session and rewrite our Discussion section by discussing some questions following reviewer #2's comments, which include *"What does this new technique brings to the knowledge of LaTex hypoxia? How does it compare with earlier models? How is this useful to managers? What are the caveats and limitations? what are the future developments? How is this technique portable to other systems?"*

Conclusions:
I think that the paper would benefit from a more comprehensive conclusion that reiterated some of the broader implications and benefits that could come from this hybrid ensemble approach. The final two sentences are really just devoted to saying that this is the first of its kind, which again reinforces some of the issues I mention in the introduction related to this being a pure modeling exercise.
Authors' Response: We will provide more discussion of the implications of this study and emphasize it in the Conclusion section.

Specific Comments

Line 15: It may benefit the reader to include a percentage value in comparison to the low RMSE value of 3204 square kilometers, which may be quite large in other coastal systems.
Authors' Response: We will add a percentage difference to illustrate the model performance in addition to RMSE and R^2.

Line 20: Suggest removing the words "by far". Because this model is the first to do this, the modifier "by far" suggests that no other groups are anywhere near this operational capability. I'm not sure if this is the intent, maybe this is meant instead to say that this ensemble model has the highest performance skill "by far".
Authors' Response: We intend to say that this ensemble model has the highest performance skill "by far". We will rewrite this sentence to avoid ambiguity, like by removing the phrase "by far".

Line 25: Suggest changing to "shelf-wide" here and elsewhere in the paper
Authors' Response: We will correct it accordingly.

Line 30: I've seen "destruction" of hypoxia used more often than "deconstruction" in the literature, suggest making this change
Authors' Response: We will correct it accordingly.

Line 41-43: Awkward phrasing, cut out "however" from sentence
Authors' Response: We will correct it accordingly.

Line 46-47: Suggest rephrasing as "An additional Bayesian model applied to summer bottom DO predictions accounts for May total nitrogen…"
Authors' Response: We will correct it accordingly.

Line 49-52: Suggest rewording as "Mechanistic prediction methods have also been applied by Laurent and Fennel (2019) to develop a weighted mean forecast that is calibrated using May nitrate loads and three-dimensional hindcast simulations over the period 1985-2018. Once calibrated, the model only requires May nitrate loads as an input to produce the seasonal forecast for a given year."
Authors' Response: We will correct it accordingly.

Line 55: Suggest changing "shortages" to "drawbacks"
Authors' Response: We will correct it accordingly.

Line 55-59: Remove periods before points 2 and 3, otherwise you can remove the colon and break them all up into single sentences. Point 2 could also be reworded slightly, reads awkwardly now. Change "year-to-year" to "interannual"
Authors' Response: We will correct it accordingly.

Line 61-62: Suggest rewording to something like "Here we aimed to provide a new

technique in HA prediction that considers both stratification and biochemical effects, and accurately produces daily forecasts of HA based on selected predictors' own forecasts."
Authors' Response: We will correct it accordingly.

Line 65-67: Hypoxic volume really hasn't been mentioned up to this point in the manuscript, and here you say that it will be neglected because HA is a better predictor anyway. Would suggest removing these sentences altogether.
Authors' Response: We will remove these sentences accordingly.

Line 71-77: I understand that some of the data used for model evaluation are described in the companion paper, but this section seems to be much more focused on derived model inputs (e.g. reanalyses and model outputs). Suggest changing the title of this section to reflect this better.
Authors' Response: We will change the title of this section to something like "Data preparation".

Line 87: Suggest changing to "… the amount of energy per volume required to homogenize the entire water column"
Authors' Response: We will correct the sentence accordingly.

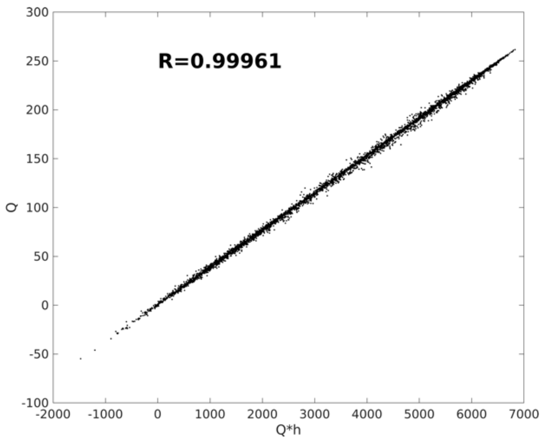
Line 95: Change "… are other two factors influencing" to "are two other factors that influence"
Authors' Response: We will correct the sentence accordingly.

Line 95-96: Could be worth mentioning that the effect of tidal mixing on stratification is neglected in this study site, since it's included as an additional term in the Simpson 1981 paper.

Authors' Response: Yes indeed. The Simpson 1981 paper did consider a tidal mixing term which is ignored in our study. We will add a sentence "The effects of tidal mixing which was considered in Simpson's (1981) equation was neglected in our study due to the relatively weaker tidal effects on stratification in the shelf when compared to the effects of river and wind".

Line 98: The first term on the right-hand side of this equation is negative in Simpson et al. (1978), but it seems like the way that this has been defined (reversing the position of water density and depth-integrated water density), that this may actually be referencing the equation of Simpson 1981. Equation 1 in Simpson 1981 also does not have "h" in the first right-hand side term, but I'm unsure if this is an error on Simpson's part since it appears in the 1978 paper. Suggest changing the reference and/or modifying the equation (may be easier just to change the reference rather than redo calculations/figures).

Authors' Response: The "h" term in the first right-hand-side term should not be there if following Simpson's (1981) work. The potential energy anomaly in Simpson (1981) is a depth-averaged term while that in the Simpson et al. (1978) paper is not. We follow the potential energy anomaly equation in Simpson (1981). We will remove the "h" in our equation and redo our calculations and figures. We think the results would not be significantly changed due to the depth range in the shelf region is not quite large. The correlation of Q and Q*h is high as 0.99961 as shown below.

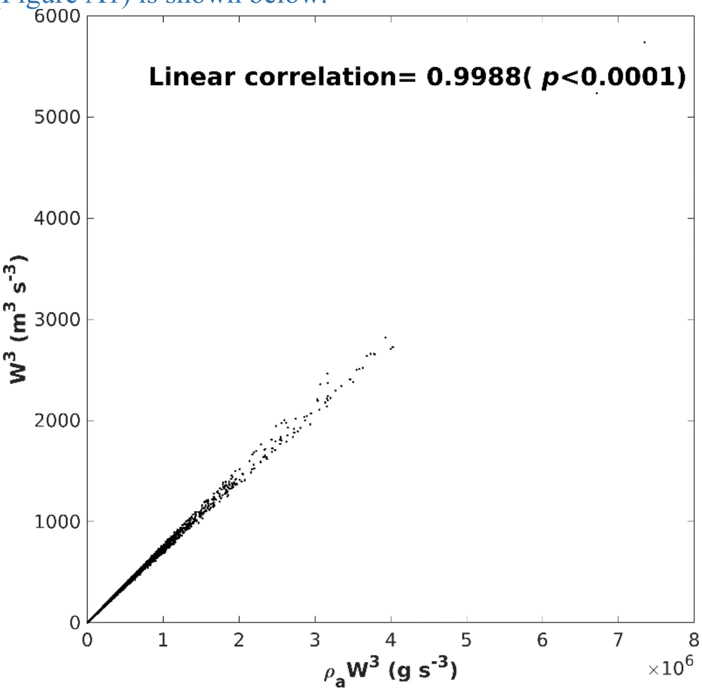


Line 110-111: Suggest referencing figure 1a here as was done in lines 90-92.
Authors' Response: We will add the referencing figure 1a here.

Line 126: Suggest changing "… estimated for the following" to "estimated by"
Authors' Response: We will correct this sentence accordingly.

Line 128: I am having trouble understanding why this equation does not match what is shown in equation 2.27 of Monteith and Unsworth (2014). It looks as if some simplification occurred such that the denominator of the exponential (T-T', where T'=36K in Monteith and Unsworth) was incorporated into the numerator in the manuscript. However, when I plot the two curves against each other I find that they are unequal, and the gap increases with increasing temperatures. At 20 degrees C, for example, this is equal to vapor pressure difference of approximately 23 Pa. Is this a relatively minor difference, or is this likely to strongly affect the correlation found when combined with W^3?

Authors' Response: Thanks for pointing it out. We did miss a T' in the numerator of the exponential term. We will correct the equation in the revision. We double-check the relationship of W^3 and the corrected ρ_a and found the strong linear relationship still holds. The corrected figure (Figure A1) is shown below.



Line 142-143: Here I would also suggest pointing the reader to figure 1a as was done in lines 90-92.
Authors' Response: Figure reference will be added here.

Line 145-146: Suggest changing phrasing to "However, global forecast model systems like HYCOM do not currently include biochemical fields."
Authors' Response: We will rewrite this sentence accordingly.

Line 156: Suggest removing this sentence and adding the correlation metric to the sentence that describes it first from lines 153-155. This earlier sentence could then read "… calculated as 19 days (R^2=0.8157, Figure A2a)."
Authors' Response: We will adjust these sentences accordingly.

Line 158: Is there a reference for this decomposition rate coefficient, or has this described in more detail in the companion manuscript?
Authors' Responses: The SOC is modeled in the accompany paper (in Eq. (8) and Eq. (10)) proportional to sedimental organic matter concentration (estimated as sedimental particulate organic nitrogen, PON_{sed}, and is output from the 3-D coupled model in the accompany paper) and a temperature-dependent decomposition rate:

$$SOC = PON_{sed} \cdot VP2N_0 \cdot e^{K_{P2N} \cdot T_b}$$

where $VP2N_0$ is a constant representing the decomposition rates of PON_{sed} at 0 ºC, K_{P2N} a constant (0.0693 ºC^{-1}) indicating temperature coefficients for decomposition of PON_{sed}, and T_b the bottom water temperature. In this study, we use the variation of Mississippi River inorganic nitrogen loads with some leading days to mimic the variation of PON_{sed} and keep the temperature-dependent decomposition rate same as that in the 3-D coupled model. Such decomposition rate follows the Q10 assumption (van't Hoff, 1898) that the reaction rate, R, depends exponentially on temperature, i.e.,

$$R = R_0 \cdot Q_{10}^{(T-T_0)/10}$$

For most biological systems, Q10 is from 2 to 3 (Bryan et al., 2008). Here, we assume it as a constant 2. R_0 is the reaction rate at temperature T_0 (measured in ºC). The SOC scheme we applied takes the R_0 as $VP2N_0$ and T_0 as 0 ºC. Thus, the above equation can be simplified as:

$$R = R_0 \cdot 2^{\frac{T_b}{10}} \approx R_0 \cdot e^{0.0693 \cdot T_b}$$

van't Hoff, J. H. (1898). Lectures in theoretical and physical chemistry: Part I: Chemical dynamics LK (p.256). London: Edward Arnold. https://UOLibraries.on.worldcat.org/oclc/220605730

Reyes, B. A., Pendergast, J. S., & Yamazaki, S. (2008). Mammalian peripheral circadian oscillators are temperature compensated. Journal of biological rhythms, 23(1), 95–98. https://doi.org/10.1177/0748730407311855

Line 163-165: I would suggest immediately describing these variables as PEA_heat, PEA_wind, and DCP_temp, rather than defining them here again.
Authors' Response: These variables have been described as they are derived in sections 2.1.1 and section 2.1.2. We thus will rewrite these two sentences to avoid prolixity.

Line 166: Can you better define what it means when you state that "multicollinearity may become a problem"? Maybe adding a short technical detail on the ramifications of this would be helpful to the reader.

Authors' Response: The multicollinearity indicates correlation among independent variables. When the multicollinearity problem occurs (i.e., strong correlations among independent variables are found), the assumption of independent variables is weakened or even collapses. It would lead to unreasonable coefficients for some highly correlated "independent" variables even though the response is well fitted by regression models. Thus, we should avoid this problem or make it less apparent. We did a best-subset searching for predictors and finally found three predictors (PEA, SOCalt, and DCP_{Temp}) that provided the best performance when applied to the GLMs and GAMs. The variance inflation factors (VIFs) of these predictors are 2.60, 2.43, and 1.23, respectively, which is less than 5 suggesting the violation of multicollinearity is negligible.

We will add the above description in our manuscript for further illustration.

Line 169-170: Are all the grid cells the same size for this model domain? Is this described in more detail in the companion paper?

Authors' Response: The sizes of the grid cells are not the same but are nearly a constant of 25.56 ± 0.17 km^2 (mean$\pm$1std). The minimum and maximum sizes are 25.18 km^2 and 25.96 km^2, respectively.

Line 188: Change "rest" to "remaining"

Authors' Response: We will change it accordingly.

Line 190-191: Change "is chosen randomly" to "are chosen randomly" and "is grouped into" to "are grouped into"

Authors' Response: We will change it accordingly.

Line 192: Suggest changing to "split at intervals of 5000 km^2"

Authors' Response: We will change it accordingly.

Line 272: Some awkward phrasing "… which impose more threatens to the shelf ecosystem."

Authors' Response: We wanted to emphasize that it is more important to increase the model performance in the hypoxic area peak during which the shelf ecosystem would face more threats than during the mild hypoxic events. We will rewrite this sentence accordingly.

Line 299: Misspelling of "procedure"

Authors' Response: We will correct it.

Line 332-333: Suggest change to "… tends to underestimate HA peak estimates (like those seen at samples 310 and 920)"

Authors' Response: We will correct it accordingly.

Line 351-352: What daily data are referred to here, the outputs derived from HYCOM or the nitrate and nitrite loadings from USGS?

Authors' Response: The daily data here are the HyCOM data and USGS nitrate and nitrite loads. We will rewrite this sentence to avoid misunderstanding.

Line 378-381: These two sentences are a bit repetitive and could be combined. I'm also not entirely clear about whether HYCOM is expected to integrate USGS runoff in the future. Is the use of daily estimates part of long-term plans for HYCOM simulations?

Authors' Response: We will simplify these two sentences. We are not sure if HyCOM modeling groups have such a plan for their global products.

Line 399: Some awkward phrasing, suggest changing to "… HA forecast capable of explaining up to 80% of the total variability"

Authors' Response: We will correct it accordingly.

Line 404: "… on HYCOM,s"

Authors' Response: It is a typo and should be "HyCOM". We will correct it accordingly.