Responses to RC1

Review of "Long-term Strengthening of the CO_2 Sink and Spatiotemporal pCO_2 dynamics in the northern Gulf of Mexico: Insights from a 22-year Satellite-based Machine Learning Reconstruction" by Jiang et al.

General Comment:

1. This manuscript applies a seasonally specific random forest model to a long term pCO_2 dataset to provide a 22 year simulation of variations in pCO_2 in the northern Gulf of Mexico. Their findings support the view that the northern Gulf of Mexico is a net sink for atmospheric CO_2, driven to a large extent by the river-influenced region. The authors also examined both thermal and non-thermal effects on pCO_2. The manuscript is well-written and provides a comprehensive literature review of prior related work in the northern Gulf. Methods are clearly described. The effort builds on and goes beyond prior work in developing a more extensive time-series and examining controlling factors on pCO_2. The authors provide an insightful analysis of the drivers of pCO_2 variations. Findings were related to seasonal and long-term SST variations, and they also identify important trends in changes in biological drivers (e.g., increasing chlorophyll) and physical factors mediated by vertical mixing and plume transport.

Response: We sincerely thank the reviewer for the positive assessment of our manuscript, including the clarity of the writing, the comprehensive literature review, the methodological description, and the insightful analysis of pCO_2 drivers.

2. It would be useful if the authors also discussed how their CO_2 flux rates compared their rates to those of prior work such as that cited in the paragraphs of the introduction on lines 69-101. Perhaps a table comparing rates from different studies could be included?

Response: We agree with the reviewer that a quantitative comparison of our CO_2 flux estimates with previous studies would strengthen the manuscript. In the revised version, we will add a new supplementary table that compares our CO_2 flux rates with those from key prior studies, including Huang et al. (2015) based on 13 cruises (2003–2012), Lohrenz et al. (2018) using a satellite-based regression tree approach (2006–2010); Kealoha et al. (2020) from SOCAT-based flux estimates; and Gomez et al. (2020) using a high-resolution biogeochemical model. We will also add a discussion paragraph to interpret the differences among these estimates in terms of methodological differences, temporal coverage, and spatial resolution. We believe this addition will provide readers with a clear context for evaluating our flux estimates.

3. Overall, this is a useful and important contribution and, provided the authors can address this and additional minor comments below, I recommend publication.

Response: We sincerely thank the reviewer for the positive evaluation and the recommendation for publication.

Specific comments:

1. Lines 106-108: Explain how transitions between the seasonal submodels occur. How to avoid abrupt transitions in the simulations?

Response: We thank the reviewer for raising this important point. In our model development, we actually tested two training strategies: a single full-season model applied to all months, and the season-specific random forest model with four sub-models (Table S1). Both approaches yielded consistent results for the subsequent seasonal and decadal trend analyses. However, we found that the season-specific model consistently outperformed the full-season model in terms of agreement with *in situ* observations, which is not surprising given that the spatiotemporal distribution and dominant controlling mechanisms of surface pCO_2 differ markedly among seasons in the nGOM. Regarding the transition between seasonal sub-models, we also tested using both adjacent seasonal sub-models for the transitional months between seasons (e.g., May–June, August–September, November–December), and found that the differences between the resulting predictions were smaller than the overall uncertainty of the model (~27.6 μatm). Therefore, no significant abrupt transitions were introduced at the seasonal boundaries. We have added a brief explanation of this transition strategy in the revised Methods section to clarify the procedure for readers.

2. Line 132: The 'g' in 'adg' stands for 'Gelbstoff', which a German word for 'yellow substance', and refers to the dissolved organic matter fraction of absorption. So the term 'adg' refers to the dissolved-detrital fraction of absorption. Please correct.

Response: Thank you for pointing out the correct definition of adg. We have revised the relevant contents throughout the manuscript following standard marine optical terminology. In Section 2.1 describing satellite parameters, the text has been corrected as: *"Satellite parameters used in this study (Fig. 1c) include sea surface temperature (SST_{sat}), salinity (SSS_{sat}), chlorophyll a (Chl-a_{sat}), and the dissolved-detrital fraction of absorption (adg)"*.

3. Lines 156-157: The in situ pCO2 data was randomly divided into three subsets: training, internal validation and external validation. How representative was the external validation subset to different regions of the Gulf? It would also be helpful to have more explanation of how the internal validation subset was used for the hyperparameter tuning, and how it usage differed from the external subset as the results for the two were very similar.

Response: We thank the reviewer for this careful and constructive question. In the revised manuscript, we have significantly expanded the description of our modeling workflow in Section 2.2 to clarify the distinct roles of the internal and external validation subsets, as well as the rationale behind our data splitting strategy.

Specifically, we now explicitly state that the internal validation subset (17.5% of the matched data) was used actively for hyperparameter tuning. We trained multiple random forest models on the training subset by varying key parameters, including the number of trees, maximum tree depth, and the number of candidate features for node splitting, and selected the optimal combination based on minimizing the prediction error on the internal validation set. Because this subset directly influenced the selection of hyperparameters, its performance carries an optimistic bias and cannot serve as an unbiased estimate of the model's generalization capability. In contrast, the external validation subset (30% of the data) was kept completely isolated throughout the entire model development process. It was evaluated only once after the final model was fully trained, providing a truly independent assessment of model performance.

Regarding the representativeness of the external validation set across different sub-regions of the nGOM, we acknowledge that this is an important concern. We adopted a global random shuffling and splitting scheme, which maximizes the use of all available samples ensuring that samples from each region are statistically balanced across the training and validation sets through repeated randomizations. In addition, we carried out independent verification of the model results against *in situ* underway measurements from individual cruises (Figure S1). In the revised supplement, we will include a new table summarizing the statistical results for each individual cruise shown in Figure S1, including the root mean square error (RMSE), coefficient of determination (R^2), and mean bias. This will allow readers to assess the model's performance across different seasons and spatial regions more transparently.

4. Lines 161-172: As mentioned above, how are the transitions between seasons managed in order to avoid discontinuities in the time-series?

Response: please see our response to Specific Comment 1 above.

5. Figure 7: This figure is especially small and it would be helpful if the plots and text could be enlarged slightly.

Response: We thank the reviewer for this suggestion. We agree that Figure 7 was too small and the text difficult to read. In the revised manuscript, we have enlarged this figure and increased the font sizes for all labels, axes, and legends. We have also applied similar improvements to Figure 5, which suffered from the same issue. These adjustments have been made to enhance overall clarity and readability, making the figures more accessible to readers.

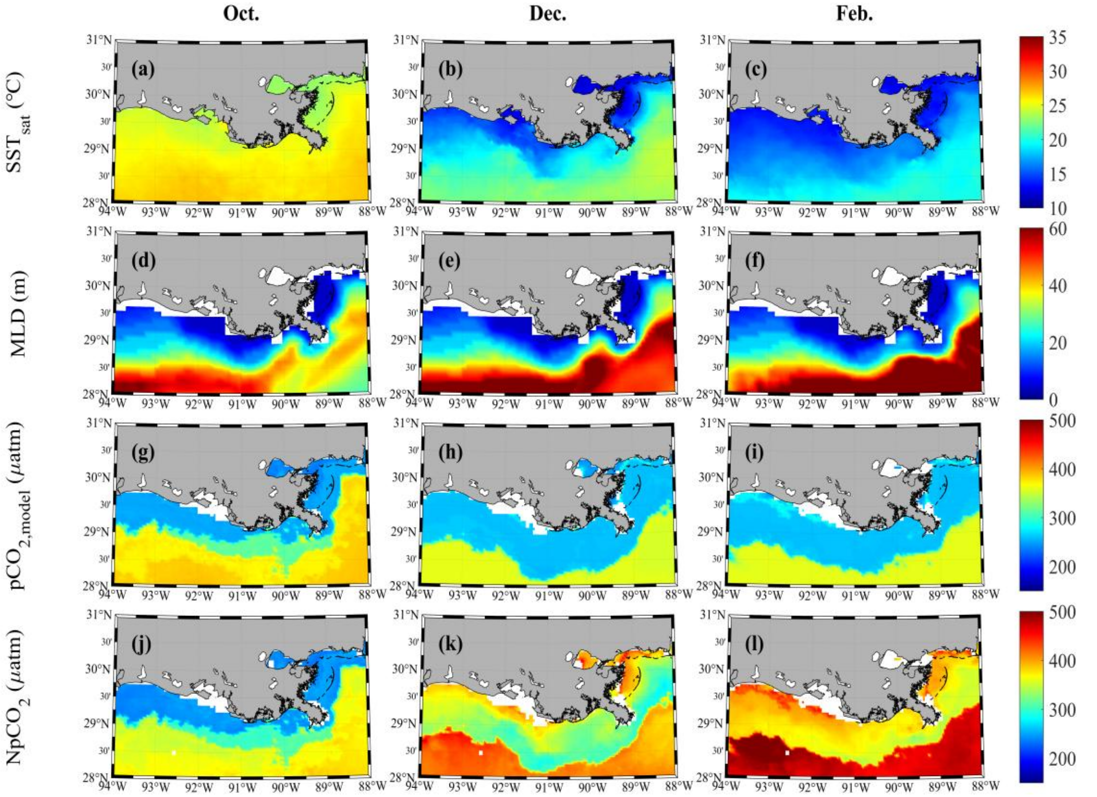

Figure 7. The autumn-winter climatology of (a, b, c) satellite-derived sea surface temperature (SST_{sat}); (d, e, f) mixed layer depth (MLD); (g, h, i) sea surface pCO_2 from the season-specific random forest model ($pCO_{2,model}$); (j, k, l) non-thermal pCO_2 ($NpCO_2$) in October (the first column), December (the second column), and February (the third column).

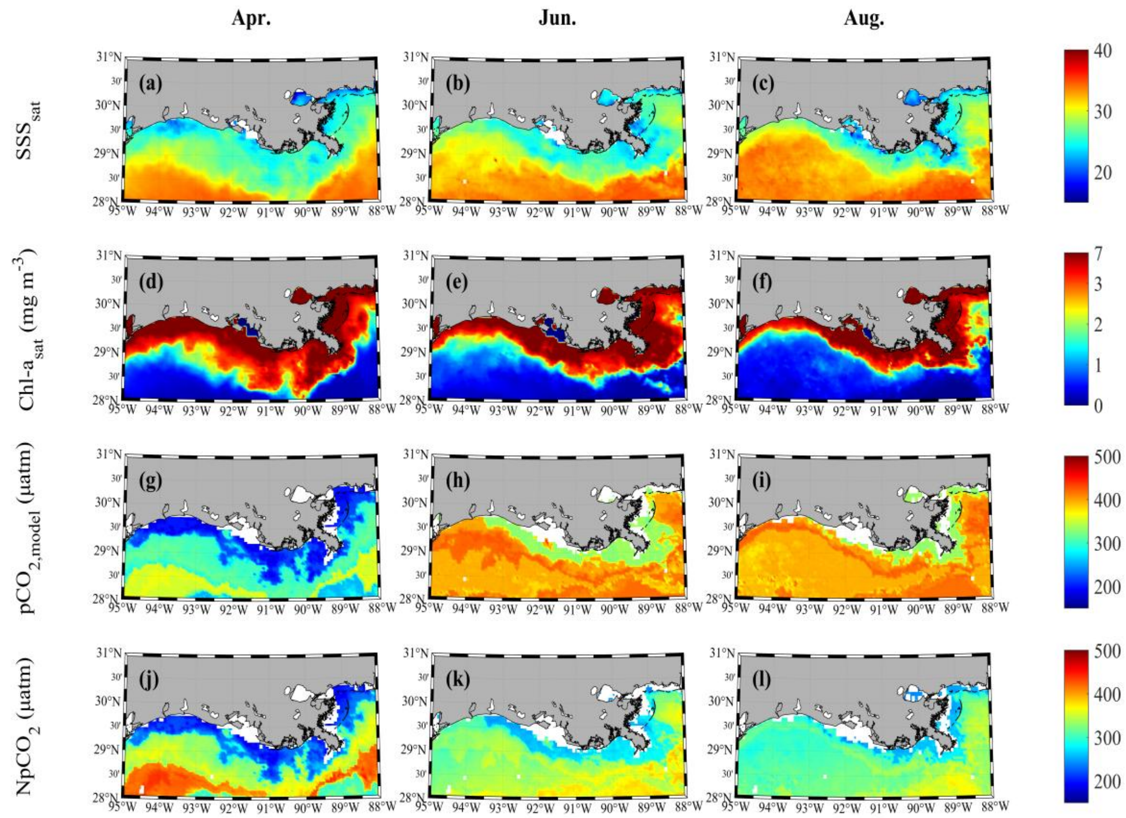

Figure 5. The spring-summer climatology of (a, b, c) satellite-derived sea surface salinity (SSS_{sat}); (d, e, f) satellite-derived chlorophyll a (Chl-a_{sat}); (g, h, i) sea surface pCO_2 from the season-specific random forest model ($pCO_{2,model}$); (j, k, l) non-thermal pCO_2 ($NpCO_2$) in April (the first column), June (the second column), and August (the third column).

6. Figure 9: Text is very difficult to read in this figure, particularly where it overlaps with data plots.

Response: We have redesigned this figure by repositioning the legend and text annotations to avoid overlap with the plotted data. We have also improved the contrast between lines and markers to improve visual clarity.

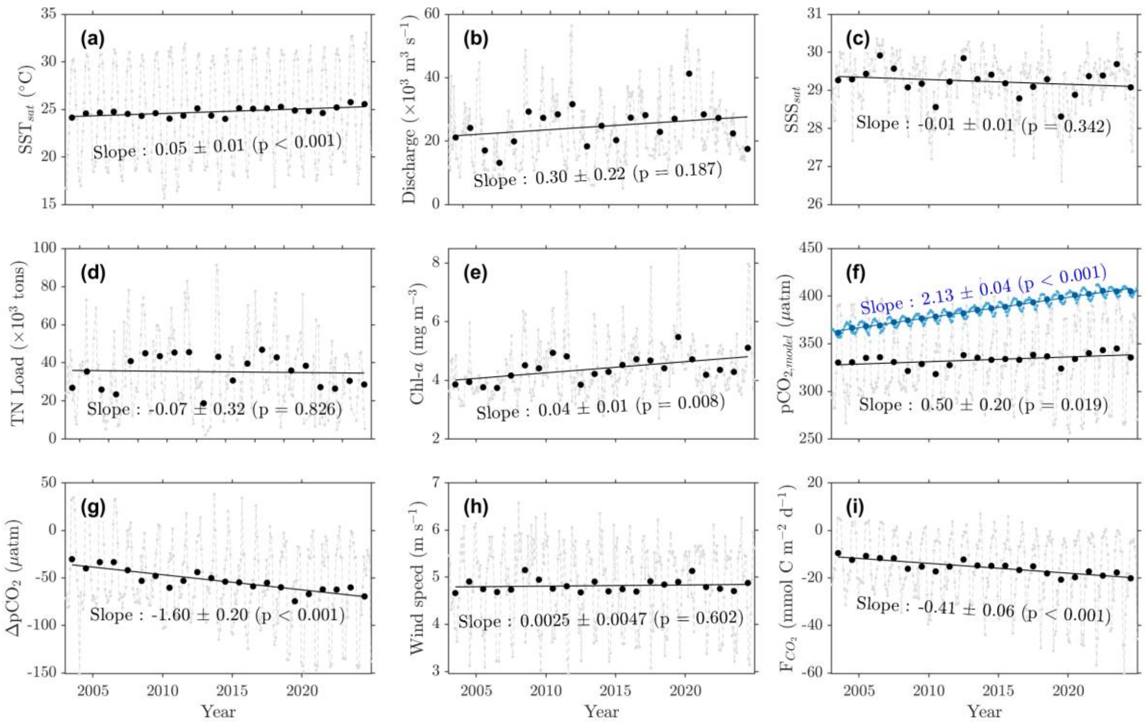

Figure 9. Long-term trends of (a) satellite-derived sea surface temperature (SST$_{sat}$); (b) Mississippi River discharge; (c) satellite-derived sea surface salinity (SSS$_{sat}$); (d) Mississippi River total nitrogen (TN) load; (e) satellite-derived chlorophyll a (Chl-a$_{sat}$); (f) sea surface pCO_2 ($pCO_{2,model}$) and atmospheric pCO_2 ($pCO_{2,air}$); (g) air-sea pCO_2 difference (DpCO_2, seawater minus atmosphere); (h) wind speed; (i) sea-to-air CO_2 flux (F_{CO2}), where negative values refer to oceanic CO_2 uptake. Gray and black dots represent monthly and annual averages, respectively, for the northern Gulf of Mexico from 2003 to 2024. The solid black lines are linear regressions fitted to the annual averages.