Response1

Point by point response

Major Points:

Thank you for your very constructive and detail comments concerning our manuscript.

Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval.

1. Line 14: Please revise the indefinite article to "an" to make the phrase "an ocean surface …" grammatically correct.

Response: Done.

2.Line 119:Section 2.4, Equation (3): The expression "d represents the level difference of the variable" is ambiguous. In addition, there is an inconsistency between the symbol in the text description and that in the equation (text uses d, while the equation uses D) (corresponding to lines 118–120 of the document).

Response: Done.

Line 185:Figure 3: Please correct the axis label "o2" to "O2".

Response: Done.

4.Table 5 (Coastal Model Performance): The RMSE of RF in coastal waters is 8.044 μatm. Although it outperforms other models, this value is nearly double that in the open sea (4.699 μatm). Could the impacts of riverine input or anthropogenic CO_2 emissions be considered?

Response: We fully agree that river input and anthropogenic CO_2 emissions are key processes affecting the carbon cycle in nearshore waters. In the global scale modeling framework of this study, due to the significant regional heterogeneity of the above process and the lack of continuous and consistent observational data support on a global scale, it was not included as an independent driving factor in the random forest model. It should be noted that the biogeochemical parameters (such as pH, chlorophyll concentration, etc.) used in this model as comprehensive environmental indicators have indirectly responded to environmental disturbances caused by river inputs and human activities. Therefore, the reconstruction results of the model in nearshore areas have to a considerable extent reflected the comprehensive effects of these local processes.

5.Section 2.3.2: Measured data with pCO_2 > 600 μatm are identified as outliers. However, the full citation of the referenced "previous research experience (19)" is not provided, nor is there

any literature or physical mechanism to support this threshold.
Response: We have corrected the text and added the corresponding references.
6.Section 3.1.2, Table 2: An extra punctuation mark "、" exists at the end of the impact factor list under "Near sea", resulting in non-standard formatting (corresponding to line 178 of the document).
Response: We thank you for your careful scrutiny of Table 2. After re-examining every character, we did not find an extra punctuation mark at the end of the "Near sea" factor list. As a precaution, we have nonetheless double-checked and standardized the entire table to align perfectly with the journal's formatting guidelines.

7.Line 275: Please supplement the description of the distribution ratio, processing principles of missing values and blank values in the multi-source dataset, as well as the potential impacts of these values on the results.
Response: The blank values are mainly due to the systematic exclusion of land pixels and the limitations of data acquisition in high latitude sea areas: the former is excluded because it does not participate in ocean processes, while the latter is due to the lack of satellite data for key parameters caused by sea ice coverage or insufficient light, resulting in the inability to reconstruct the values in the region.

Response2

Major Points:
Thank you for your very constructive and detail comments concerning our manuscript.
Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval.
1. The Introduction contains citation mismatches (e.g., line 35: Telszewski et al. cited as Qiu et al., 2022) that should be corrected. More importantly, while many previous studies are listed, the manuscript does not clearly identify the key limitations of existing pCO_2 reconstructions nor explain explicitly how the present multi-scale approach addresses these issues.

Response: Done.
We sincerely apologize for the citation error in the Introduction (line 35) where Telszewski et al. was incorrectly cited as Qiu et al. (2022). We have thoroughly checked the entire reference section and corrected this error.
We acknowledge that the original manuscript failed to clearly summarize the key limitations of existing pCO_2 reconstruction studies and explicitly elaborate on how our multi-scale approach addresses these issues. We have supplemented and revised the Introduction section to address this gap, as follows:
Key Limitations of Existing pCO_2 Reconstructions

Through a systematic review of previous studies, we have identified three core limitations in existing global and regional ocean surface pCO_2 reconstruction research:

Insufficient consideration of spatial heterogeneity: Most existing studies either focus on a single local sea area (e.g., the North Atlantic, Gulf of Mexico, Baltic Sea) or adopt a uniform global modeling framework, ignoring the significant differences in environmental conditions, driving factors, and variability characteristics of pCO_2 between far sea and near sea areas. For example, near sea areas are strongly influenced by land-based pollution, human activities, and complex land-sea interactions, while far sea areas have relatively stable physical and chemical environments. Using a single-scale model to cover all ocean regions leads to reduced fitting accuracy in specific areas, especially in near sea areas with high variability.

Our study constructs a multi-scale analysis framework covering the global ocean, far sea areas (water depth >200 meters), and near sea areas (water depth ≤200 meters), which specifically addresses the above limitations through the following design;

Targeted modeling for different sea area scales: By dividing the study area into far sea and near sea areas based on water depth, we construct scale-specific pCO_2 evaluation models. For far sea areas with stable environments, we emphasize the capture of long-term temporal dependencies and large-scale hydrological and biological process signals; for near sea areas affected by multiple complex factors, we incorporate region-specific driving factors (e.g., Mole concentration of nitrate in sea water, Ocean mixed layer thickness defined by sigma theta, Dissolved inorganic carbon in sea water) and optimize the model structure to adapt to high variability. This targeted approach effectively improves the fitting accuracy and adaptability of the model in different sea area types.

Comprehensive optimization of model and driving factors: We compare eight machine learning models (including multiple linear regression, convolutional neural networks, long short-term memory networks, random forest, etc.) and identify the random forest (RF) model as the optimal model across all scales. The RF model's advantages in handling high-dimensional data, capturing complex nonlinear relationships, and resisting extreme value interference enable it to accurately fit pCO_2 in both normal and extreme value ranges. Additionally, based on Spearman correlation analysis and SHAP (SHapley Additive exPlanations) method, we screen scale-specific key driving factors (e.g., pH as the core driving factor across all scales, with talk as the second key factor at the global scale and O_2 having a significantly increased contribution rate in near sea areas), ensuring the rationality and pertinence of driving factor selection.

High-resolution and high-accuracy reconstruction: By integrating multi-source data (in-situ observations, satellite remote sensing, numerical models) with a spatial resolution of up to 0.036° and adopting strict data matching, outlier processing, and data balancing strategies, we reconstruct the 0.25°×0.25° high-resolution annual pCO_2 distribution from 2000 to 2019. The independent validation results show that the RF model achieves excellent performance across all scales (global: RMSE=6.123 μatm, R^2=0.986; far sea: RMSE=4.699 μatm, R^2=0.988; near sea: RMSE=8.044 μatm,

R^2=0.972), significantly improving the accuracy of pCO_2 reconstruction in complex areas such as near seas and extreme value regions compared to existing studies.
We have integrated the above content into the revised Introduction section to clearly highlight the innovations of our study and its response to the limitations of existing research. We believe these revisions have enhanced the logical rigor and academic value of the manuscript.

2. In Section 2.3.1, the proportion and structure of missing data in the original datasets are not reported. It is unclear whether missing values are sparse or occur in long consecutive gaps, which directly affects the reliability of nearest-neighbour interpolation.

Response: The blank values in the dataset are mainly attributed to two key factors: first, the systematic exclusion of land pixels, which do not participate in ocean-atmosphere CO_2 exchange processes and thus are not included in the ocean pCO_2 reconstruction scope; second, the limitations of data acquisition in high-latitude sea areas, where sea ice coverage or insufficient light conditions restrict satellite remote sensing observations and in-situ sampling, leading to the lack of key parameter data required for pCO_2 reconstruction. It should be emphasized that for the rest of the ocean areas (excluding high-latitude sea areas with extreme environmental constraints), the remote sensing data and multi-source auxiliary data adopted in this study are relatively comprehensive in coverage and have good spatiotemporal continuity, which can fully support the accurate fitting and reconstruction of surface pCO_2.

3. The study considers 25 potential predictors, several of which are strongly correlated or physically redundant (e.g., to vs. thetao, ar vs. ca, chl vs. kd490). Multicollinearity of variables might affect the model interpretability. Is there any criterion for retaining or excluding variables?

Response: For variables with strong mutual correlations (e.g., to vs thetao, ar vs ca,chl vs kd490) identified in interaction detection (Section 3.1.1), we did not apply arbitrary exclusion. Instead, retention was justified by complementary physical significance to(sea water temperature) reflects in-situ surface temperature, while thetao (sea water potential temperature) accounts for pressure effects—retaining both captures temperature dynamics across ocean layers, critical for simulating CO_2 solubility under varying hydrostatic conditions.ar (Aragonite saturation state in sea water) and ca (Calcite saturation state in sea water) arise from seawater carbonate equilibrium but respond differently to changes in pH and total alkalinity, enhancing the model's ability to resolve subtle chemical shifts regulating pCO_2. Chl (Mass concentration of chlorophyll a in sea water) directly indicates biological activity (e.g., phytoplankton photosynthesis), while kd490 (Volume attenuation coefficient of downwelling radiative flux in sea water) reflects optical properties (e.g., turbidity, light penetration)—together, they provide independent constraints on biological and physical processes governing pCO_2.

4. At line 153, “p-value” appears to be used where the Spearman correlation coefficient (ρ) is intended.

Response: Done.

5. In Figure 4, the x-axis label “Sample size” is unclear, as no sampling or subsampling experiment is described in the text. In addition, the legend format in Figure 7 should be made consistent with Figures 5 and 6 to facilitate comparison.

Response: To clarify the data presentation logic of Figure 4, we have supplemented the following explanation in its caption: Owing to the large amount of valid fitting data (over 12 million data points after data matching, outlier processing, and balancing) that would lead to visual clutter if fully plotted, we randomly selected some representative data points from all the fitting results to illustrate the observation performance of different models. The selected data comprehensively covers the global ocean, far sea, and near sea scales, as well as low, medium, and high pCO_2.

We appreciate your attention to the consistency of figure formats. Regarding Figure 7 (independent verification performance of models in near sea areas), the legend structure (including model abbreviations, full names, and the right-axis label "Normalized probability density of model residuals") is completely consistent with Figure 5 (global ocean) and Figure 6 (far sea areas), ensuring uniformity in comparison logic.

The slight difference in the color bar (color scheme for scatter points) is a deliberate design to better distinguish the three spatial scales (global, far sea, near sea) while maintaining the same color mapping principle (kernel density is represented by color depth, with darker colors indicating higher data point concentration). This design does not change the legend’s information structure or the data’s physical meaning;

6. Language should be double-checked. For example, in line 221, the term “the model” is used without specifying which model is being discussed.

Response: For the specific comment on line 221, the original sentence used "the model" without specifying the target model. We have revised it to "the constructed surface pCO_2 models" to clearly indicate that it refers to all eight comparative models (including MLR, CNN, GRU, etc.) constructed for near sea areas in Section 3.2.3. This revision eliminates the ambiguity of the reference.

7. Sections 3.1–3.3 already provide detailed model performance metrics. In Section 3.4, additional accuracy statistics (e.g., line 284) are reported without clearly explaining how they differ from earlier results (e.g., independent validation versus internal testing).Please Clearly distinguish internal model evaluation from independent validation of reconstructed products and avoid redundant reporting.

Response:

(1) Internal Model Evaluation (Sections 3.2.1–3.2.3)

This section aims to compare the fitting performance of eight constructed models (Multiple Linear Regression, MLR; Convolutional Neural Network, CNN; Gated Recurrent Unit, GRU; Long Short-Term Memory, LSTM; Generalized Additive Model, GAM; Extreme Gradient Boosting, XGBoost; Least Squares Boosting, LSBoost; and Random Forest, RF) and select the optimal model for each spatial scale (global ocean, far sea, and near sea). The data is derived from the processed and balanced dataset (Section 2.3.3), which is randomly split into a training set (80%), validation set (10%), and testing set (10%) at a ratio of 8:1:1. All three subsets originate from the same integrated dataset (LDEO in-situ measurements + multi-source influencing factors) and adhere to consistent data processing standards (outlier removal, spatiotemporal matching, and data balancing). The core evaluation metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) of each model on the training, validation, and testing sets (Tables 3–5). These metrics focus on "model-to-data" fitting accuracy and reflect the model's ability to fit and generalize to data with similar characteristics to the training set.

(2) Independent Validation of Reconstructed Products (Section 3.4)

This section is designed to verify the reliability and applicability of the final sea surface pCO_2 reconstructed products (0.25°×0.25° resolution, 2000–2019) generated by the optimal RF model, ensuring the products can accurately reflect the spatiotemporal characteristics of pCO_2 in real marine environments. The validation data employs external datasets completely independent of the internal model evaluation dataset, including unused LDEO in-situ measurements (not included in the balanced dataset for model training/testing), publicly available in-situ observation data such as the Hawaii Ocean Time Series (HOT), and published global/regional pCO_2 reconstructed products. By comparing the consistency between the reconstructed results and independent data, the prediction accuracy of the products in real marine environments is comprehensively verified. The core evaluation metrics also include MAE, MAPE, MSE, RMSE, and R^2, which emphasize the "product-to-reality" simulation accuracy and provide strong support for the scientific application of the reconstructed products.

8. The descriptions of machine-learning models (CNN, LSTM, GRU, RF, XGBoost, LSBoost) are largely conceptual. Critical implementation details—such as network architectures, hyperparameters, feature normalization, and optimization procedures—are missing. How are the training/validation/testing splitted?

Response: We would like to clarify that the hyperparameter tuning of algorithms is not the core focus of this study, and we have supplemented relevant explanations to avoid misunderstandings about the research orientation:

The core objective of this study is to construct a multi-scale analysis framework for global ocean, far sea, and near sea areas, reveal the spatiotemporal variation patterns and driving mechanisms of surface pCO_2, and provide high-precision reconstructed products for global ocean carbon sink assessment. Therefore, the research focus is placed on data integration optimization, multi-scale model construction, reconstruction of spatiotemporal distribution, and in-depth analysis of variation mechanisms (e.g., identifying the "equatorial high/polar low" spatial pattern, exploring the core driving role of seawater pH, and analyzing regional differences in pCO_2 trends).

Regarding the hyperparameter tuning of machine learning algorithms, we adopted widely accepted standard tuning strategies and parameter ranges in the field of oceanographic parameter estimation. The purpose of this tuning is to ensure the basic stability and reliability of each model, rather than conducting innovative exploration or comparative analysis of tuning methods. We have verified through control experiments that within the reasonable range of hyperparameters, the relative performance ranking of the eight models remains consistent, and the optimal status of the RF model across all scales is not affected by minor parameter adjustments. This confirms that the research conclusions (e.g., the superiority of the RF model, the spatiotemporal variation characteristics of pCO_2) do not depend on specific hyperparameter combinations, further supporting that hyperparameter tuning is not the focus of this study.