Response to review of **“Evaluation of CMIP6 Models Performance in Simulating Historical Biogeochemistry across Southern South China Sea”**

Manuscript egusphere-2024-72

Response to Anonymous Referees:

We sincerely appreciate the time and effort invested by both the reviewers and the editor in evaluating our paper titled **"Evaluation of CMIP6 Models Performance in Simulating Historical Biogeochemistry across Southern South China Sea"** submitted for publication in Biogeosciences. We are grateful for the positive feedback and the insightful comments provided, which is detailed in this report and also in the upcoming revised manuscript. The majority of the suggestions put forth by the reviewers have been incorporated, and in the limited cases where we have not, we have provided a detailed description of the justification for each decision. For ease of reference, we have provided a detailed point-by-point response to the reviewer’s comments with line numbers in the response representing the changes in revised manuscript.

*This report contains point-by-point responses to each comment from both reviewers.

REVIEWER-1 MAJOR COMMENTS

Comment 1:

An in-depth process evaluation is needed

First, irrespective from the valuable effort (and outcome) of this work, that is evaluated marine biogeochemistry models embedded in CMIP6 ESMs, I think further attention to other physical drivers or control variables would be needed to better understand the performance of the models.

Another need would be to scrutinize inter-parameters relationship (chl-biomass, biomass-nitrate). I think the reader could be interested in further understanding of biases propagation across the marine biogeochemical cycles.

A similar attention would be needed also to dive a bit further into the model process parameterization. For instance, not all model simulates prognostically chlorophyll. This latter is derived from the carbon-to-chlorophyll ratio and phytoplankton biomass (see Séférian et al. 2020, who did an in-depth evaluation of model parameterization).

Response:

- Thank you for highlighting the importance of analysing the physical drivers or control variables of models to enhance our understanding of their performance. In this paper, our primary focus is to assess the performance of CMIP6 models in reproducing historical biogeochemical variables. While we appreciate the suggestion to include analysis of physical drivers or control variables, we believe that expanding the scope in this direction may detract from the main focus of our research. Moreover, it's worth noting that previous studies such as Jin et al. (2023) and Fan & Zhou (2023) have already evaluated the performance of CMIP6 models on some physical drivers/control variables (i.e., SST, Asian-Pacific Oscillation and precipitation), including those relevant to our study domain (southern South China Sea) and common models utilized in our research (ACCESS-ESM1-5, CanESM-5, GFDL-ESM4, MIROC, MPI-ESM-1-2-HAM, MPI-ESM-1-2-HR, MPI-ESM-1-2-LR, NorESM2-LM, NorESM2-MM & MRI-ESM2-0). However, we acknowledge the value of conducting a separate analysis on the performance of individual model's physical drivers or control variables alongside biogeochemical responses in future research. This suggestion has been duly noted and included as a recommendation in our paper **(L575)**.
- Thank you for highlighting the importance of inter-parameters relationship. We also agree with your suggestion, and as it aligns with the scope of our study, we accordingly produced the correlation between Chlorophyll-Phytoplankton biomass **(L401)** and Nitrate-Phytoplankton biomass **(L431)**. As a result, our analysis revealed strong positive correlations between chlorophyll concentration and phytoplankton biomass in most CMIP6 models, indicating robust primary production and nutrient availability. However, we also observed deviations in some models, likely stemming from differences in parameterizations and structural complexities. These findings highlight the necessity of considering model uncertainties in our interpretations. Similarly, our examination of nitrate-phytoplankton relationships showed significant positive correlations in some CMIP6 models, while others displayed weaker correlations. This discrepancy potentially due to variations in biogeochemical tracers within model frameworks could influence model efficacy. The relevance of these discrepancies to their model dynamics is discussed in detail in **L421 to L460**.
- Thank you for bringing up the importance of model process parameterization. We acknowledge your point that not all models incorporate chlorophyll prognostically. However, delving deeply into model process parameterization might extend beyond the scope of our present study, which primarily aims at evaluating model reproducing skill on biogeochemical variables. Nevertheless, we recognize the potential value in comparing the parameterization processes of different models for future research. We suggest considering this as a separate study to explore the performance of these models more comprehensively in the future. Additionally, we have included this suggestion in our recommendations **(L575 – L582)**.

****

Comment 2:

The choice of biogeochemical tracers

The authors focused on key biogeochemical variables: chlorophyll, plankton biomass, nitrate and oxygen. All of these variables are analyzed at surface oceans — which makes sense for biological markers such as chlorophyll and plankton biomass but a surface analysis can hide model biases for oxygen and nitrate (oxycline and nutricline). There an expanded analysis including nitrate and oxygen profiles across the multi-model ensemble would be relevant.

I also think that phosphate and net marine productivity would be relevant to liase with global studies such as Kwiatkowski et al. 2020.

Response:

- Thank you for bringing up the importance of analysing the bias at depth to understand the oxycline and nutricline dynamics of the models. In response, considering the complex bathymetry of southern SCS region, we have addressed this concern by presenting the spatial distribution of seasonal variations in nitrate and oxygen at two distinct depths (70m and 1000m) for each model, rather than providing profiles **(L277 – L325)**. We have discussed the biases observed in each model and the ensemble, noting consistent negative bias in nitrate concentrations across models, with MPI-based models showing the least positive bias at 70m but shifting towards a negative bias at 1000m. Oxygen biases followed a similar trend, with UKESM1-0-LL displaying significant negative bias at 70m and positive bias at 1000m. Additionally, CanESM5 and MIROC-based models exhibited high negative biases at 1000m. These biases could potentially arise from inaccuracies in parameterizing relevant processes, exclusion of critical processes and structural uncertainties in model formulation. These discrepancies were discussed in detail at **L283 – L293** for nitrate and **L304 – L325** for oxygen. The depth of 70 meters has been selected to depict the dynamics of the nutricline/oxycline in the shelf break region, while a depth of 1000 meters has been chosen to represent the deep layer.
- We sincerely thank the reviewer for recommending the inclusion of additional biogeochemical parameters such as phosphate and net marine productivity, following the work of Kwiatkowski et al. 2020. However, after careful consideration, we choose to include only variables that were consistently available across all selected models' historical and projection scenarios. Since phosphate and net productivity variables were absent in some of the historical and projection scenarios, we prioritized models that shared common variables in both their historical and projection scenarios.

****

Comment 3:

Choice of reference datasets

Finally I’m also concerned by the use of the single CMEMS data are taken as ground truth observations although they aren’t.

As far as I am aware, CMEMS is a model-based data product. For the ocean physics the data product benefit from observation assimilation (it is thus a reanalysis) but for marine biogeochemistry it is only the marine biogeochemical model PISCES without any constraints.

I understand a focus on a given regional domain could pose challenge in terms of data availability. However, there are many high-resolution datasets for surface chlorophyll, net primary productivity etc. based on remote-sensing observation that can provide support to this work.

I would recommend to use them in addition to CMEMS. Indeed, past work (Lee et al. 2016) showed that this data product do not outcomes standard CMIP5 models when compared to true insitu observations.

Response:

- Thank you for your concern about the usage of reference data (CMEMS). Unfortunately, CMEMS is the only available timeseries hindcast data for the biogeochemistry in southern South China Sea region. As we stated already in the manuscript **(L137)** that CMEMS biogeochemistry product quality has been validated by Mercator-Ocean and they have confirmed and published the quality of this data through comparisons with recognized datasets like Ocean Color, World Ocean Atlas and Globcolour products in their Quality Information Document (QuID; Perruche et al., 2019).
- In order to improve more confidence of this dataset in southern South China Sea region, we have discussed some literatures in our revised manuscript **(L142 – L151)**, in which, authors have validated this data product with the in-situ measurement in this study region, i.e., Wahyudi et al, (2023) validated the POC, Chlorophyll, Dissolved Oxygen, Nitrate, Phosphate and Silicate obtained from CMEMS biogeochemistry product by comparing it with in-situ data collected during the Widya Nusantara Expedition 2015 (Triana et al., 2021) in the upwelling area of southwestern Sumatra waters. They found that the mean absolute percentage error values were lower than 15%, indicating the reliability of the CMEMS biogeochemistry model data in our study area. Additionally, Chen et al, (2023) also used the daily chlorophyll concentration data from the same CMEMS biogeochemical product in south china sea region. By utilizing this CMEMS biogeochemistry model dataset, Wahyudi et al. (2023) and Chen et al. (2023) highlights the proficiency of the CMEMS biogeochemistry model data in reproducing both the climatic patterns and fluctuations observed within its biogeochemical variables in southern South China Sea. This gave us confidence in utilizing the CMEMS biogeochemical dataset as the reference model to assess other models in this region (southern South China Sea).
- Thank you for recommending the other available dataset. Although there are other datasets available for surface chlorophyll and net primary productivity, the purpose of using the single reference dataset (CMEMS) is to ensure consistency in the evaluation process. By relying on a homogeneous dataset, we aim to enhance the reliability of our evaluation results and greater confidence in the findings.

****

REVIEWER-1 MINOR COMMENTS

Comment 1:

L9 why? they are poorly constrained

Response:
Thank you for bringing our oversight on explain this matter. The emphasis on selecting the biogeochemical variables is not primarily on constraint, but rather on highlighting the significance of key biogeochemical tracers and their availability across the chosen models and their corresponding projection scenarios. Accordingly, we have also explained this matter in our revised manuscript **(L08)**.

****

Comment 2:

L12 overestimations or underestimations, in what? (quantitative measures or score would be useful)

Response:
Thank you for bringing up this important point. We apologize for the oversight in not clarifying. The overestimations or underestimations refer to quantitative measures. We have incorporated this clarification into the revised manuscript as per your recommendation **(L12)**.

****

Comment 3:

L23-25 This statement is inexact, recent work from Kwiatkowski et al 2020 shows that marine NPP in 2100 remains largely uncertain

Response:

We acknowledge the inexactness of the statement regarding marine NPP uncertainty in 2100. Upon thorough examination of Kwiatkowski et al.'s (2020) findings, we recognize that the uncertainty in primary production largely persists, even within the CMIP6 framework. To address this, we have diligently incorporated the relevant insights from Kwiatkowski et al. (2020) into our revised manuscript **(L24 – L30)**, thereby ensuring the accuracy of our discussion on this matter.

****

Comment 4:

L47 Consider to refer to Séférian et al. 2020 for the evaluation of global ESMs

Response:

Thank you for your valuable suggestion to consider Séférian et al. (2020). After careful review, we agree that their work on global ESMs evaluation is relevant to our study. Thus, we have incorporated a citation to Séférian et al. (2020) in our revised manuscript **(L56)**.

****

Comment 5:

L58 This aspect has to be analysed with caveats because for several bgc models chlorophyl is derived from phytoplankton biomass improving linkage with zooplankton response (see major comments)

Response:

We sincerely appreciate your input on this matter. However, in this particular context, our aim was to convey the findings of Petrik et al. (2022), who extensively examined the model's simulation of zooplankton and delved into the relationship between zooplankton and chlorophyll-a. It is worth noting, as we have highlighted in our revised manuscript **(L65)**, that Petrik et al. (2022) underscored "*chlorophyll-a as a proxy for phytoplankton"* for several reasons: (a) it provides the most comprehensive observations across both time and space, (b) it is more readily observable compared to phytoplankton biomass, which typically requires physical sampling, or primary production, which relies on physical experiments or algorithms based on satellite chlorophyll data that yield varied outcomes, and few other reasons.

****

Comment 6:

L102 it would be nice to have a table of the references datasets used here
L112 CMEMS : it might be good to make it clear if CMEMS data is reanalysis and model reconstruction.
As far as I am aware of CMEMS relies on NEMO-PISCES+ assimilation for ocean hydrodynamics. PISCES, the marine bgc model adjust to improved physics but it is free (no assimilation). Therefore using CMEMS for evaluating marine bgc models is comparing models with another (more constrained) model.

Response:

We sincerely appreciate your thoughtful consideration of this matter and your concern. The utilization of the CMEMS biogeochemical dataset in our study is primarily due to its status as the only available timeseries hindcast biogeochemistry dataset for our study domain, specifically the southern South China Sea region. Furthermore, we have referenced literature that has validated this data product for this region through comparisons with in-situ measurements, instilling confidence in its use for evaluating the models in this area. We kindly wish to inform you that a detailed response addressing this specific issue was provided in our response to **Reviewer-1 major comment 3**.

****

Comment 7:

L125 Table 1: I'm suprised to see PISCES (IPSL, CNRM, EC-Earth) and MARBL (CESM2) excluded from this exhaustive analysis — what are the reason?

Response:
We appreciate your attention to this issue. As explained in **L132** that our model selection procedure was based on the availability of selected biogeochemical variables across historical or projected scenarios. Consequently, the EC-Earth models do not include the Phytoplankton biomass (phyc) variable, while CESM2 lacks the dissolved Oxygen (o2) variable in its historical dataset. Additionally, the CNRM and IPSL-based models showed a standard deviation >50 mg/m3 for the chlorophyll variable compared to reference data, resulting in their exclusion from the analysis.

****

Comment 8:

L129 Ranking: no cross-variable evaluation? SST-Chlorophyll, etc. (see major comments)

Response:
We appreciate your emphasis on evaluating the model's biogeochemistry in relation to its physical drivers. However, the primary focus of our paper is the assessment of CMIP6 biogeochemical model outcomes. While we acknowledge the significance of examining biomass correlations with physical drivers, delving into this aspect might deviate from the central goal of our paper. Instead, we have thoroughly analysed the model's performance concerning Chlorophyll-Phytoplankton biomass and Nitrate-Phytoplankton biomass correlations. We kindly wish to highlight that we addressed this concern in detail in our response to **Reviewer-1 major comment 1**.

****

Comment 9:

L159-160: please consider sharing scripts and data for this work (EGU journals recommandations)

Response:
Thank you for your thoughtful consideration of our work and for your interest in the data. We have indeed provided the source link to the dataset we utilized **(L156)** from CMEMS biogeochemistry hindcast dataset (ID: GLOBAL_MULTIYEAR_BGC_001_029) and CMIP6 dataset which was obtained from ESGF data portal, which are freely accessible for public use. Furthermore, we employed widely accepted basic formulae for our statistical analyses, as detailed in our manuscript **(L73 – L81)**.

****

Comment 10:

L164-165 yes and no there are several paper that indicates how seasonal cycle can help to constrain projections (Behrenfeld et al. 2006, Kwiatkovski et al. 2017, etc..)

Response:
Thank you for bringing this matter to our attention. We sincerely apologize for any confusion caused by our previous statement. Upon careful review of the works by Behrenfeld et al. (2006) and Kwiatkovski et al. (2017), we have revised our statement accordingly. Instead of asserting that *"the yearly cycle of seasons does not fully capture the long-term changes associated with climate change,"* we have amended it to reflect that *"the yearly cycle of seasons partially captures the long-term changes associated with climate change"*. These long-term changes encompass shifts in average temperatures, alterations in precipitation patterns, changes in the frequency and intensity of extreme weather events, and other systemic shifts that extend beyond the periodicity of seasonal cycles. While temporal cycles are indeed important components of climate variability, they offer only a partial perspective on the broader and more profound changes occurring in the Earth's climate system. Thus, our revised statement in **L198 – L202** aims to convey that "*the yearly cycle of seasons partially captures the long-term changes associated with climate change*".

****

Comment 11:

L183: typo: ensemble mean ?
'Moderate' needs to be quantified with skillscore

Response:
Thank you for pointing out this oversight. We have corrected the typo to *"ensemble"* in the manuscript. Furthermore, we've also provided clarification on the term "*moderate"* by delineating its specific meaning in context. In the discussion of spatial variation and bias in **L220**, *"moderate"* denotes the model's performance at an average level in simulating the variables quantitatively, relative to the reference data. Likewise, in reference to Taylor's diagram in **L481**, *"moderate"* indicates the model's placement between the last and first models in TD.

****

Comment 12:

Figures 2-5: hard to see model differences.
I would present model baises against the reference

Response:
We sincerely appreciate your insightful feedback. We acknowledge your concern regarding the clarity of the model differences in Figures 2-5 in old manuscript. In response to your suggestion, we have changed the seasonal climatology figures to illustrate the seasonal bias against the reference data in our revised manuscript **(Figs. 2–25)**.

****

Comment 13:

Figure 7: why for surface oxygen of MPI-based models outbest the others whereas it is not the case for the other variables ? For these later MIROC-ES2L outbests the other models.
These differences needs to be explained. My guess is that SST biaises in MPI-based models are much lower than the other models in this zone which explains why surface oxygen is better represented.

Response:
Thank you for bringing to our knowledge about the differences in model out best behaviour and also acknowledge for providing us the solution as well. Following your suggestion, we have conducted a thorough review of some literature. Our findings align with your suggestion and discussed in our revised manuscript **(L535 – L542)** that MPI-based models exhibit superior performance in simulating oxygen levels in our study, largely attributed to their effective physical drivers. Research conducted by Jin et al. (2023) verifies this by demonstrating that MPI-based models excel in replicating SST patterns, particularly during boreal winter and summer. They are noted for their minimal SST biases, particularly in Asian marginal seas, which contributes to their accuracy in representing surface oxygen variables in our study.

****

REVIEWER-2 GENERAL COMMENTS

Comment 1A:

The aim of the paper is to rank 13 CMIP6 ESM simulations based on their ability to reproduce selected observed biogeochemical variables. However, the dataset that the author chose is not strictly observations. Based on the link they provided in line 123, the CMEMS ocean biogeochemistry product is based on the PISCES model output (although it is forced with reanalysis product). I also noticed that among the 13 CMIP6 ESMs, the authors have not chosen IPSL-CM6A ESM, which includes PISCES as its ocean biogeochemical model. I understand that in-situ observations may be rare in this region, but to truly assess the CMIP6 ensemble and individual models, I suggest the authors could compare the CMIP6 models with satellite-derived chlorophyll-a and primary production, as well as the World Ocean Atlas product for nitrate and oxygen.

Response:

- Thank you for your concern about the usage of reference data (CMEMS). Unfortunately, CMEMS is the only available timeseries hindcast data for the biogeochemistry in southern South China Sea region. As we stated already in the manuscript **(L137)** that CMEMS biogeochemistry product quality has been validated by Mercator-Ocean and they have confirmed and published the quality of this data through comparisons with recognized datasets like Ocean Color, World Ocean Atlas and Globcolour products in their Quality Information Document (QuID; Perruche et al., 2019).
- In order to improve more confidence of this dataset in southern South China Sea region, we have discussed some literatures in our revised manuscript **(L142 – L151)**, in which, authors have validated this data product with the in-situ measurement in this study region, i.e., Wahyudi et al, (2023) validated the POC, Chlorophyll, Dissolved Oxygen, Nitrate, Phosphate and Silicate obtained from CMEMS biogeochemistry product by comparing it with in-situ data collected during the Widya Nusantara Expedition 2015 (Triana et al., 2021) in the upwelling area of southwestern Sumatra waters. They found that the mean absolute percentage error values were lower than 15%, indicating the reliability of the CMEMS biogeochemistry model data in our study area. Additionally, Chen et al, (2023) also used the daily chlorophyll concentration data from the same CMEMS biogeochemical product in south china sea region. By utilizing this CMEMS biogeochemistry model dataset, Wahyudi et al. (2023) and Chen et al. (2023) highlights the proficiency of the CMEMS biogeochemistry model data in reproducing both the climatic patterns and fluctuations observed within its biogeochemical variables in southern South China Sea. This gave us confidence in utilizing the CMEMS biogeochemical dataset as the reference model to assess other models in this region (southern South China Sea).
- While we appreciate your suggestion of alternative datasets, our decision to utilize CMEMS as the sole reference dataset was made to maintain consistency in the evaluation process. By adhering to a homogeneous dataset, we aim to ensure the integrity and reliability of our evaluation results, thereby instilling greater confidence in our findings.
- Additionally, IPSL-based models showed a standard deviation >50 mg/m^3 for the chlorophyll variable compared to reference data, resulting in their exclusion from the analysis.

Comment 1B:

Since the paper also looks at the seasonal trend of biogeochemical properties, it could benefit from exploring whether different CMIP6 models can capture phytoplankton phenology (e.g., Racault et al., 2015; Gittings et al., 2018), which is an important indicator.

Response:

- We appreciate your insightful comment and apologize for any confusion regarding our approach. Our examination primarily focused on the seasonal spatial climatology (now presented as seasonal spatial bias in **Figs. 2-25**), not the seasonal trend map.
- While we acknowledge the significance of exploring phytoplankton phenology in CMIP6 models, as suggested by the references you provided, conducting such studies requires extensive time and resources and also, we afraid that incorporating phenology in this study could potentially diverge the main scope of our current investigation. However, we recognize the importance of this aspect and have duly noted it as a potential avenue for future research in our study **(L459 – L463)**. Thank you for bringing this to our attention.

****

Comment 2:

Indeed, most of the biological activity occurs near the surface layers of the ocean, but it's important to consider the biogeochemical dynamics near the seabed, particularly in shelf seas, as they can have complex structures through interactions of ocean physics with biological processes, such as export and remineralization. I would appreciate the inclusion of depth profiles and benthic concentrations of oxygen and nitrate – this would provide a more thorough assessment of the biogeochemical properties. Furthermore, most of the biogeochemical models used in CMIP6 are not specifically built for shelf seas. It would be interesting to see whether these models can represent nutrient and oxygen distribution at shallower depths.

Response:

- Thank you for bringing up the importance of analysing the bias at depth to understand the oxycline and nutricline dynamics of the models. In response, considering the complex bathymetry of southern SCS region, we have addressed this concern by presenting the spatial distribution of seasonal variations in nitrate and oxygen at two distinct depths (70m and 1000m) for each model, rather than providing profiles **(L277 – L325)**. We have discussed the biases observed in each model and the ensemble, noting consistent negative bias in nitrate concentrations across models, with MPI-based models showing the least positive bias at 70m but shifting towards a negative bias at 1000m. Oxygen biases followed a similar trend, with UKESM1-0-LL displaying significant negative bias at 70m and positive bias at 1000m. Additionally, CanESM5 and MIROC-based models exhibited high negative biases at 1000m. These biases could potentially arise from inaccuracies in parameterizing relevant processes, exclusion of critical processes and structural uncertainties in model formulation. These discrepancies were discussed in detail at **L283 – L293** for nitrate and **L304 – L325** for oxygen. The depth of 70 meters has been selected to depict the dynamics of the nutricline/oxycline in the shelf break region, while a depth of 1000 meters has been chosen to represent the deep layer.

****

Comment 3:

Although the authors put a great effort in evaluating CMIP6 model outputs, the model structures could also be evaluated; how biogeochemical tracers are represented, and whether these representations affect the performance of the model in the southern SCS. Perhaps the authors can add another table which biogeochemical tracers these models represent (e.g., in MEDUSA-2 (UKESM), it does not represent diazotrophic phytoplankton, explicitly calculates phytoplankton chlorophyll, and uses N as model currency, while in OECO-2 (MIROC), it has diazotrophic phytoplankton with C as model currency and includes Phosphate as nutrients), and perhaps also how they are formulated, especially when it involves trophic transfer (e.g. nutrient uptake, zooplankton grazing, and phytoplankton growth, and plankton mortality). These additions can add some discussion on how model representation (and structure) may affect model performance in the shelf seas, instead of repeatedly saying that underestimation/overestimation is due to zooplankton grazing/phytoplankton productivity/nutrient uptake.

Response:

We sincerely appreciate your insightful comments regarding the model structures. Following your suggestion, we have incorporated an overview of how tracers and model structure affect performance into our discussion, specifically when examining inter-parameter relationships such as chlorophyll-biomass and biomass-nitrate. We explained that most models, except for UKESM1-0-LL and MIROC-based models, use carbon as the primary measure of phytoplankton biomass and include nitrate and phosphate to constrain growth rates, resulting in a weaker correlation with nitrate. The MPI-ESM1-2-LR model has a new nitrogen-fixing formulation that better captures the nitrogen response. The UKESM1-0-LL model uses nitrogen as its primary measure, leading to a stronger response to nitrate. The MIROC-ES2L model includes the phosphorus cycle to strongly depict the phosphorus limitation on diazotrophic phytoplankton growth, which could explain the negative correlation with nitrate. For the GFDL-ESM4 model, the negative correlation could be due to its categorization of phytoplankton by size and nutrient content, attributing specific N:P ratios. This summarized version is discussed in detail at **L437 – L459** in revised manuscript.

****

Comment 4:

The presentation of the results can also be improved. I think it will be easier to follow the results if the authors describe the observed distribution of nitrate, chlorophyll, phytoplankton biomass, and oxygen, then compare them with the model. For the figures, it would be more interesting to see the difference between the CMEMS data and CMIP6 outputs with better figure resolution (especially figure 6). Additional discussion on regions where bias usually occurs in different models will also be interesting (e.g., the shelf seas between Sumatra, the Malaysian peninsula, and Borneo are always high in phytoplankton biomass for UKESM, CanESM5, ACCESS, MPI-ESM1-2, NorESM2).

Response:

- We sincerely appreciate your insightful feedback and acknowledge your concern regarding the clarity of the results. Following your suggestion, we have modified the seasonal climatology figure in our revised manuscript to better illustrate the seasonal bias against the reference data in **Figs. 2 – 25**.
- Additionally, we have incorporated your recommendation to discuss the spatial diversity in model bias. Our revised manuscript now elaborates this in **L236 – L251**.

****

REVIEWER-2 SPECIFIC COMMENTS

Comment 1:

L12 – perhaps the authors can add a % or number on the degrees of overestimation and underestimation.

Response:
Thank you for bringing up this important point. We apologize for the oversight in not clarifying. The overestimations or underestimations refer to quantitative measures. We have incorporated this clarification into the revised manuscript as per your recommendation **(L12)**.

****

Comment 2:

L22-23 - Based on CMIP6 models, NPP trend is uncertain, apart maybe at the Southern Ocean (Tagliabue et al., 2021)

Response:
Thank you for your valuable feedback. We acknowledge the inaccuracy in our original statement regarding marine NPP uncertainty in 2100. After thoroughly reviewing the findings of Tagliabue et al. (2021) and Kwiatkowski et al. (2020), we recognize that significant uncertainty in primary production indeed persists, even within the CMIP6 framework. To address this, we have incorporated the insights from Kwiatkowski et al. (2020) and Tagliabue et al. (2021) into our revised manuscript **(L24 – L30)** to ensure a accurate discussion on this matter.

****

Comment 3:

L33-34 - This is not always the case - OBGC models can give the seemingly good representation of historical climate pattern but for the wrong reason. Furthermore, OBGC model results is dependent on its physical forcings (see Sinha et al., 2010)

Response:
Thank you for bringing this matter to our attention. In response to your suggestion, we have gone through some literatures and made changes accordingly in **L36 – L43**.

****

Comment 4:

L60 – typo: Tjiputra et **al**, (2020)

Response:
Thank you for pointing out this oversight. We have corrected accordingly in **L69**

****

Comment 5:

L61-72 – I'm not so sure if these are appropriate examples. Maybe add studies like Kwiatkowski et al., 2020, Hinrichs et al., 2023

Response:
Thank you for your valuable suggestion and for providing literature on this topic. We have thoroughly reviewed the references you recommended, and we have incorporated relevant findings from Kwiatkowski et al., 2020 into our manuscript **L79**.

****

Comment 6:

L83-L85 – Why only phytoplankton, chlorophyll, nitrogen, and oxygen? Why not net primary production and or carbon?

Response:
Thank you for your concern about the selected variables. We chose to focus on phytoplankton, chlorophyll, nitrogen, and oxygen for following reasons:

1. These variables are fundamental tracers for biological and nutrient dynamics in ocean systems.
2. Phytoplankton and chlorophyll variables serve as effective proxies for primary production, thus encompassing the essential aspect of net primary production.
3. Given that the southern SCS region is recognized as a typical oligotrophic area where primary productivity is primarily constrained by nutrient availability, we specifically included nitrate and oxygen.
4. We considered variables that were consistently available across all selected models' historical and projection scenarios to ensure comparability and consistency in our analysis. Even phosphate variable is unavailable in some of the selected model's scenarios. Thus, phosphate is also excluded.

****

Comment 7:

L103-L105 - This sounds like phytoplankton is controlling the physical biogeochemical process?

Response:
We sincerely apologize for the confusion caused by our statement. Upon a careful review, we rephrased the statement as **"***Within the southern SCS, extensive observations have demonstrated that phytoplankton growth, serving as the primary source of organic matter, significantly influences oceanic carbon cycles. This growth is*

influenced by monsoon-driven physical and biogeochemical processes, with phytoplankton demonstrating a notable sensitivity to these environmental dynamics." in **L125 – L128**.

****

Comment 8:

L122-123 – is this the hindcast global ocean biogeochemistry? Do you also use the GlobColour for chlorophyll? Please be more specific.

Response:

- We sincerely apologize for the confusion made. The CMEMS product used in this study is the hindcast global ocean biogeochemistry dataset, which can be found in CMEMS biogeochemistry hindcast dataset (ID: GLOBAL_MULTIYEAR_BGC_001_029) **(L156).**
- We did not use GlobColour data in our study. Instead, Mercator Ocean used GlobColour data to validate the chlorophyll data within the CMEMS hindcast biogeochemistry dataset, as mentioned in **L139 – L142**.

****

Comment 9:

L125 – Perhaps, instead of having 2/3 ESMs with the same OBGC model, maybe choose one of them instead, so you can also look at other models such as PISCESv2 (Aumont et al., 2016), MARBL (Long et al., 2021), BFM5.2 (Lovato et al., 2022)?

Response:

Thank you for suggesting a method to choose models. We will incorporate this method in our future studies. Additionally, as explained in **L132** that our current model selection procedure was based on the availability of selected biogeochemical variables across historical or projected scenarios. Based on this, CESM2 model was excluded from our study due to the absence of the dissolved oxygen (o2) variable and EC-Earth models do not include the phytoplankton biomass (phyc) variable, in their historical dataset.

****

Comment 10:

L132 – Do you mean visualised using taylor diagram? How do you calculate model/data comparison using a diagram?

Response:

We sincerely apologize for the confusion caused by our statement. Upon a careful review, we replaced the word *"calculated"* to *"visualized"* in **L166**.

****

Comment 11:

L164-165 Can you provide a reference on this statement?

Response:

Thank you for bringing this matter to our attention. We sincerely apologize for any confusion caused by our previous statement. Upon careful review of the works by Behrenfeld et al. (2006) and Kwiatkovski et al. (2017), we have revised our statement accordingly. Instead of asserting that *"the yearly cycle of seasons does not fully capture the long-term changes associated with climate change,"* we have amended it to reflect that *"the yearly cycle of seasons partially captures the long-term changes associated with climate change"*. These long-term changes encompass shifts in average temperatures, alterations in precipitation patterns, changes in the frequency and intensity of extreme weather events, and other systemic shifts that extend beyond the periodicity of seasonal cycles. While temporal cycles are indeed important components of climate variability, they offer only a partial perspective on the broader and more profound changes occurring in the Earth's climate system. Thus, our revised statement in **L198 – L202** aims to convey that "*the yearly cycle of seasons partially captures the long-term changes associated with climate change*".

****

Comment 12:

L172 – but CMEMS data is not really observation, isn't it?

Response:
We sincerely apologize for the confusion caused by our statement. CMEMS is not observation data. We replaced the word *"observed"* to *"reference"* in our revised manuscript **L209**.

****

Comment 13:

L177-L179 - perhaps spell out how these models represent their phytoplankton growth and chlorophyll concentration? And compare it to models that have better RMSD?

Response:
Thank you for highlighting this concern. The model's representation of biological tracers was discussed in Inter-variable relations section in **L400**.

****

Comment 14:

L184, 219 – what is acceptable range?

Response:
Thank you for highlighting this concern. We have addressed it by representing the acceptable bias range based on models with small mean bias error. Accordingly, in the revised manuscript, we have specified the acceptable range as $\leq \pm 0.15$ mg/m^3 for chlorophyll in **L221** and $\leq \pm 0.4$ mmol/m^3 for phytoplankton in **L235**.

****

Comment 15:

L186 - why is UKESM not overestimating chlorophyll, but overestimates phytoplankton carbon?

Response:
Thank you for your concern regarding this matter. In **L226 – L231** we explained that *"UKESM1-0-LL model explicitly simulates chlorophyll concentrations, allowing for a more accurate representation of chlorophyll levels (Sellar et al., 2019). However, UKESM1-0-LL uses nitrogen as its primary model currency, which results in a more pronounced quantitative representation of nutrient levels. This might lead to enhanced nutrient uptake by phytoplankton due to differences in model parameterizations and consequently result in the overestimation of phytoplankton biomass."* This could explain why the model does not overestimate chlorophyll but does overestimate phytoplankton.

****

Comment 16:

L232-L242 – maybe move this to the study domain part instead of on the results section?

Response:
Thank you for your suggestion. Accordingly, we shifted this part to study domain section in **L110 – L120**.

****

Comment 17:

L251 – Can you give example of the important processes?

Response:
Thank you for your insightful comment. The important processes that may be overlooked by some ESMs include nutrient cycling, light availability, temperature variations, and phytoplankton phenology. These processes play important roles in shaping the seasonal patterns of biogeochemistry in marine ecosystems. We have clarified this point in the revised manuscript in **L395**.

****

Comment 18:

L286 **-** Why do you think this is? could it be that the ESMs in CMIP6 is developed based on the condition of the open ocean, but not the shelf seas? Or is it because the resolution is too coarse for shelf seas?

Response:
Thank you for your insightful question. Indeed, both factors you mentioned could contribute to the observed performance of the ESMs in simulating biogeochemical variables. The ESMs in CMIP6 are primarily developed based on the conditions of the open ocean, which may not fully capture the complexities of shelf seas. Additionally, the coarse resolution of these models may not adequately resolve the fine-scale processes occurring in shelf sea environments. Together, these factors likely contribute to the moderate to poor performance of the ESMs in simulating biogeochemical variables, as mentioned in **L507 – L511**.

****

Comment 19:

Figure 6 could do with higher resolution.

Response:
Thank you for your suggestion regarding Figure 6. Accordingly, we have resolved the clarity of the Figure, which is now **Fig. 28**.

****

Reference:

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., et al. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, *444*(7120), 752–755. https://doi.org/10.1038/nature05317

Chelton, D. B., Deszoeke, R. A., Schlax, M. G., El Naggar, K., & Siwertz, N. (1998). Geographical variability of the first baroclinic Rossby radius of deformation. *Journal of Physical Oceanography*, *28*(3). https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2

Friedrichs, M. A. M., Hood, R. R., & Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *53*(5–7). https://doi.org/10.1016/j.dsr2.2006.01.026

Glessmer, M. S., Oschlies, A., & Yool, A. (2008). Simulated impact of double-diffusive mixing on physical and biogeochemical upper ocean properties. *Journal of Geophysical Research: Oceans*, *113*(8). https://doi.org/10.1029/2007JC004455

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020a). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020b). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5). https://doi.org/10.5194/gmd-13-2197-2020

Jin, S., Wei, Z., Wang, D., & Xu, T. (2023). Simulated and projected SST of Asian marginal seas based on CMIP6 models. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1178974

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., et al. (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, *17*(13), 3439–3470. https://doi.org/10.5194/bg-17-3439-2020

Laufkötter, C., John, J. G., Stock, C. A., & Dunne, J. P. (2017). Temperature and oxygen dependence of the remineralization of organic matter. *Global Biogeochemical Cycles*, *31*(7). https://doi.org/10.1002/2017GB005643

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019a). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4). https://doi.org/10.1029/2018MS001400

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019b). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018MS001400

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*(1). https://doi.org/10.1002/2016MS000737

Petrik, C. M., Luo, J. Y., Heneghan, R. F., Everett, J. D., Harrison, C. S., & Richardson, A. J. (2022). Assessment and Constraint of Mesozooplankton in CMIP6 Earth System Models. *Global Biogeochemical Cycles*, *36*(11). https://doi.org/10.1029/2022GB007367

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., et al. (2020). Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports*. https://doi.org/10.1007/s40641-020-00160-0

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and Evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739

Sinha, B., Buitenhuis, E. T., Quéré, C. Le, & Anderson, T. R. (2010). Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models. *Progress in Oceanography*, *84*(3–4). https://doi.org/10.1016/j.pocean.2009.10.003

Six, K. D., & Maier-Reimer, E. (1996). Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, *10*(4). https://doi.org/10.1029/96GB02561

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO2. *Journal of Advances in Modeling Earth Systems*, *12*(10). https://doi.org/10.1029/2019MS002043

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., & Vialard, J. (2021). Persistent Uncertainties in Ocean Net Primary Production Climate Change Projections at Regional Scales Raise Challenges for Assessing Impacts on Ecosystem Services. *Frontiers in Climate*, *3*. https://doi.org/10.3389/fclim.2021.738224

Thompson, B., Tkalich, P., Malanotte-Rizzoli, P., Fricot, B., & Mas, J. (2016). Dynamical and thermodynamical analysis of the South China Sea winter cold tongue. *Climate Dynamics*, *47*(5–6). https://doi.org/10.1007/s00382-015-2924-3

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., et al. (2020). Ocean biogeochemistry in the Norwegian Earth System Model version 2 (NorESM2). *Geoscientific Model Development*, *13*(5), 2393–2431. https://doi.org/10.5194/gmd-13-2393-2020

Wang, Z., Brickman, D., Greenan, B., Christian, J., DeTracey, B., & Gilbert, D. (2024). Assessment of Ocean Temperature Trends for the Scotian Shelf and Gulf of Maine Using 22 CMIP6 Earth System Models. *Atmosphere - Ocean*, *62*(1). https://doi.org/10.1080/07055900.2023.2264832