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 Referee #1:

We sincerely appreciate the time and effort invested by both the reviewer 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 each response refers to the revised manuscript. Additionally, we have included the recently added discussions, revised figures and newly added references in this “response to the review” report for the convenience of the reviewer/editor to refer.

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 **(L564 – L577)**.

- 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 and Nitrate-Phytoplankton biomass **(L395 – L466)**. To explain this, we added the new section to the revised manuscript as follows:

  "4.2 Inter-variable Relationship
  4.2.1 Chlorophyll – Phytoplankton

  The correlation between chlorophyll and phytoplankton serves as a critical indicator of marine ecosystem health and productivity. As chlorophyll is a pigment essential for photosynthesis in phytoplankton, its concentration is often used as a proxy for phytoplankton in aquatic environments (e.g. Petrik et al., 2022). A strong positive correlation between chlorophyll and phytoplankton signifies robust primary production and nutrient availability, highlighting favorable conditions for marine life. Conversely, a weak or negative correlation may indicate nutrient limitation, environmental stressors, or other factors affecting phytoplankton growth. Understanding this correlation provides valuable insights into ecosystem dynamics, nutrient cycling and the impacts of environmental changes on marine ecosystems. However, it's important to note that not all models simulate chlorophyll concentrations prognostically. Instead, some models derive chlorophyll concentrations from the carbon-to-chlorophyll ratio and phytoplankton biomass. This approach acknowledges the intricate relationship between chlorophyll production and phytoplankton biomass, ensuring a comprehensive representation of primary productivity in marine environments. The correlation between chlorophyll and phytoplankton biomass can also provide insights into whether a model produces chlorophyll prognostically or not. If a model *simulates chlorophyll concentration prognostically, there should be a strong positive correlation between chlorophyll concentration and phytoplankton biomass. This correlation arises from the direct influence of phytoplankton biomass on chlorophyll production through photosynthesis. On the other hand, if chlorophyll concentrations are derived from the carbon-to-chlorophyll ratio and phytoplankton biomass, the correlation may still exist but could be influenced by additional factors such as nutrient availability and environmental conditions. Therefore, analysing the correlation between chlorophyll and phytoplankton biomass can help discern the modeling approach used to simulate chlorophyll dynamics within the ocean model. During our analysis of the linear regression between chlorophyll and phytoplankton, most of the selected CMIP6 models showed a strong positive correlation between chlorophyll and phytoplankton concentrations* ***(Fig. 26)****. Notably, ACCESS-ESM1-5, MIROC-ES2L, and MRI-ESM2-0 demonstrated particularly robust positive correlations, with coefficient (R) reaching 1. However, examination of the 95% confidence interval suggests that the relationship between chlorophyll and phytoplankton in MIROC-ES2H, MPI-based models and NorESM-based models deviates from proximity. This disparity in confidence intervals among those models could arise from the differences in model parameterizations and structural complexities, resulting in differing levels of uncertainty in the simulated relationships between chlorophyll and phytoplankton. Models with simpler representations of biological processes or less accurate parameterizations may exhibit wider confidence intervals due to increased uncertainty in their outputs. Specifically, MPI-based and NorESM-based models employed the HAMOCC version of the biogeochemistry model, which does not explicitly simulate chlorophyll concentrations (Paulsen et al., 2017; Tjiputra et al., 2020).*

4.2.1 Nitrate – Phytoplankton

Through the application of linear regression analysis, we investigated the correlation between surface nitrate levels and phytoplankton biomass. While analysing the various CMIP6 models, it was observed that none of the models displayed a correlation (R) > 0.8. However, among the selected models, CanESM5-ConOE, MIROC-ES2H, MPI-ESM1-2-LR, MRI-ESM2-0, and UKESM1-0-LL exhibited a statistically significant positive correlation ($R > 0.5$, $p < 0.001$) with surface nitrate and phytoplankton. Conversely, the remaining models demonstrated considerably weaker positive correlations, with only CanESM5, GFDL-ESM4, MIROC-ES2L and MPI-ESM1-2-HR displaying a slight negative correlation. This slight negative correlation could stem from various factors that it may reflect discrepancies in those model dynamics, such as the representation of nutrient uptake or phytoplankton growth rates. Biological processes within the models might not accurately capture the complexities of phytoplankton-nutrient interactions. For example, variations in biogeochemical tracers within model frameworks could influence model efficacy. Specifically, except UKESM1-0-LL and MIROC-based models, all other selected models utilize carbon as their primary model currency for representing phytoplankton biomass, incorporating explicit calculations for phytoplankton biomass and they also utilize nitrate and phosphate to constrain bulk phytoplankton growth rates alongside temperature and light. Consequently, their representation of phytoplankton biomass exhibited a weaker correlation with nitrate. Despite the use of carbon tracer, MPI-ESM1-2-LR incorporates a newly resolved nitrogen-fixing formulation within its biogeochemistry model. This updated formulation introduces an additional prognostic phytoplankton class, replacing the diagnostic formulation of nitrogen-fixation utilized in MPI-ESM-LR (Paulsen et al., 2017; Mauritsen et al., 2019). As a result, this adjustment enables the model to capture the nitrogen response to phytoplankton biomass positively. UKESM1-0-LL employed nitrogen as its primary currency, resulting in a more pronounced quantitative representation of phytoplankton biomass in response to increased nitrate levels compared to the other models ***(Fig. 27)****. While MIROC-ES2L primarily utilizes nitrogen as its tracer, it also integrates the phosphorus cycle within the model framework to accurately depict the strong phosphorus limitation on the growth of diazotrophic phytoplankton (Hajima et al., 2020). Consequently, this incorporation of the phosphorus cycle may account for phosphorus limitation, resulting in the observed negative correlation between nitrate and phytoplankton biomass within our study area. In the case of GFDL-ESM4, the negative correlation between nitrate and phytoplankton could potentially originate from their model parametrization. In their framework, phytoplankton were categorized based on size and functional type, with small phytoplankton being nitrogen-rich and large phytoplankton* *phosphate-rich, thereby attributing characteristic N:P ratios (Stock et al., 2020). Thus, differences in parameterizations, data initialization and model resolution could contribute to divergent simulated responses (Séférian et al., 2020). This discrepancy underscores the variability among model outputs and highlights the importance of further scrutinizing model performances based on their parametrization and phenological structural. It is important to analyse how these models are formulated and their roles in nutrient uptake, zooplankton grazing, phytoplankton growth, and plankton mortality within the trophic transfer processes. This approach will aim to refine our comprehension of the intricate dynamics governing marine ecosystems within each model."*

- 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 **(L L564 – L575)**.

****

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 **(L270 – L320)**. The depth of 70 meters has been selected to depict the dynamics of the nutricline/oxycline in the shelf break region of Sunda Shelf region, while a depth of 1000 meters has been chosen to represent the deep layer. Accordingly, we have discussed the biases

- For nitrate **(L278 – L291)** as follows:

  "Furthermore, delving into model biases at deeper levels, especially concerning nutrient dynamics, will provide more insights into the model's accuracy in simulating the nutricline. Consequently, we analysed the nitrate concentrations at depths of 70m ***(Figs. 11 – 13)*** *and 1000m* ***(Figs. 14 - 16)****. In contrast to surface nitrate, most models exhibited a negative bias at deeper layers (70m and 1000m), with an average range of -2 to -8 mmol/m^3 across the study area. Among these models, MPI-based models showed the least negative bias at 70m depth; however, as depth increased to 1000m, their bias shifted towards the positive* ***(Fig. 13 and 16****, respectively****)****. MIROC-based and MPI-based models exhibited the least bias in nitrate concentrations at both surface and deep layers compared to reference data. This may be attributed to the near balance achieved between nitrogen cycle sources (such as nitrogen fixation, atmospheric nitrogen deposition, and riverine nitrogen input) and sinks (including denitrification, nitrous oxide emission, and sedimentary loss) over the long spin-up period (Mauritsen et al., 2019; Hajima et al., 2020). In contrast, CanESM5-based models demonstrated minimal nitrate bias at the surface but showed varying positive and negative biases in deep layers. These discrepancies arise from the simplified parameterization of denitrification in their BGC models. In these models, denitrification in the deep*

layers is set to balance the rate of nitrogen fixation and is vertically distributed in proportion to the detrital remineralization rate. However, in reality, nitrogen fixation and denitrification are not constrained to balance within the water column at any single location; rather, denitrification primarily occurs in anoxic areas (Swart et al., 2019). Notably, no seasonal bias in all selected models were observed at the deep layer (1000m; ***Fig. 16****).”*

- For oxygen **(L293 – L313)** as follows:

  “During the observation of oxycline dynamics in the selected models, it was noted that the oxygen exhibited a positive bias at a depth of 70m, transitioning to a negative bias with increasing depth (1000m) ***(Fig. 20 and 25)****. Moreover, UKESM1-0-LL consistently exhibited a substantial positive bias from the surface to the depth of 70m (~50 mmol/m³) and shifts to negative bias of -40 mmol/m³ at 1000m relative to its surface bias. Similarly, CanESM5 and MIROC-based models displayed markedly high negative biases at a depth of 1000m, but with comparatively lesser negative biases at 70m. Multiple factors could contribute to biases in the simulation of nutricline/oxycline dynamics by models. Inaccuracies in simulating nutricline dynamics may arise from errors in parameterizing physical, chemical and biological processes relevant to these dynamics. In winter, most models overestimate oxygen levels at the surface and at a depth of 70 meters. This positive bias in oxygen concentration may result from excessively intense winter mixing of cold, oxygen-rich waters from the northern boundary of the southern SCS into the Sunda Shelf region (Thompson et al., 2016), which transports an excessive amount of surface oxygen to deeper layers. Additionally, nutrient trapping issues may also contribute to the remaining model bias (Six & Maier-Reimer, 1996). Moreover, the exclusion of relevant processes or feedback mechanisms influencing nutricline dynamics within the model, such as nutrient upwelling, microbial remineralization and ocean stratification, may lead to biased simulation outcomes. Additionally, structural uncertainties embedded in the model formulation, including simplifications or assumptions regarding complex processes, may also play a role in generating biases in simulation results. For example, advancements in model parameterization and representation of biogeochemical fluxes have led to consistent improvements in the mean states of nutrient dynamics in CMIP6 models, such as GFDL-ESM4, MIROC-based, MPI-ESM1-based, and NorESM2-based models. Specifically, improvements in GFDL-ESM4 performance are attributed to a series of updates and changes in model physics (such as mixing and climate dynamics) and biogeochemical parameterizations such as the implementation of a revised remineralization scheme for organic matter that depends on oxygen and temperature (Laufkötter et al., 2017).”*

- 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 **(L136)** 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 **(L141 – L150)**, 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 **(L13)**.

****

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 **(L23 – L29)** as follows:

"For example, Kwiatkowski et al. (2020) discovered that the multi-model global mean projections from the Coupled Model Intercomparison Project Phase 6 (CMIP6), under high-emission to low-emission scenarios, indicate a consistent decrease in net primary production. Notably, there is a significant increase in inter-model uncertainty compared to CMIP5. This increased uncertainty is linked to changes in the temporal patterns of phytoplankton resource availability and grazing pressure within CMIP6 (Kwiatkowski et al., 2020). This carries significant implications for evaluating ecosystem impacts on a regional scale. (Tagliabue et al., 2021)."

****

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 **(L53)**.

****

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 **(L61 – 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 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 **L130** 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 **(L155)** 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 **(L72 – L80)**.

****

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 **(L197 – L201)** 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 **L219**, *"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 **L477**, *"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)**. The revised figures can also be found in this report below.

****

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 and found our findings align with your suggestion, and included a statement in **L531 – L537** as follows:

"MIROC-ES2H demonstrated superior performance across all variables except oxygen, consistently achieving scores above 0.2. For oxygen, MPI-ESM1-2-LR ranked highest due to its exceptional spatial representation and accurate seasonal pattern captured in the Taylor diagram. This outperformance of MPI-based models for oxygen can be attributed to the effectiveness of the physical drivers within that model. Jin et al. (2023) showed that MPI-based models excel in simulating climatological sea surface temperature (SST) during boreal winter and summer, and are among the top performers in reproducing SST climatology in Asian marginal seas due to their minimal SST biases. Thus, MPI-based models outperform others in replicating surface oxygen variables in our study."

****

Revised Figures:

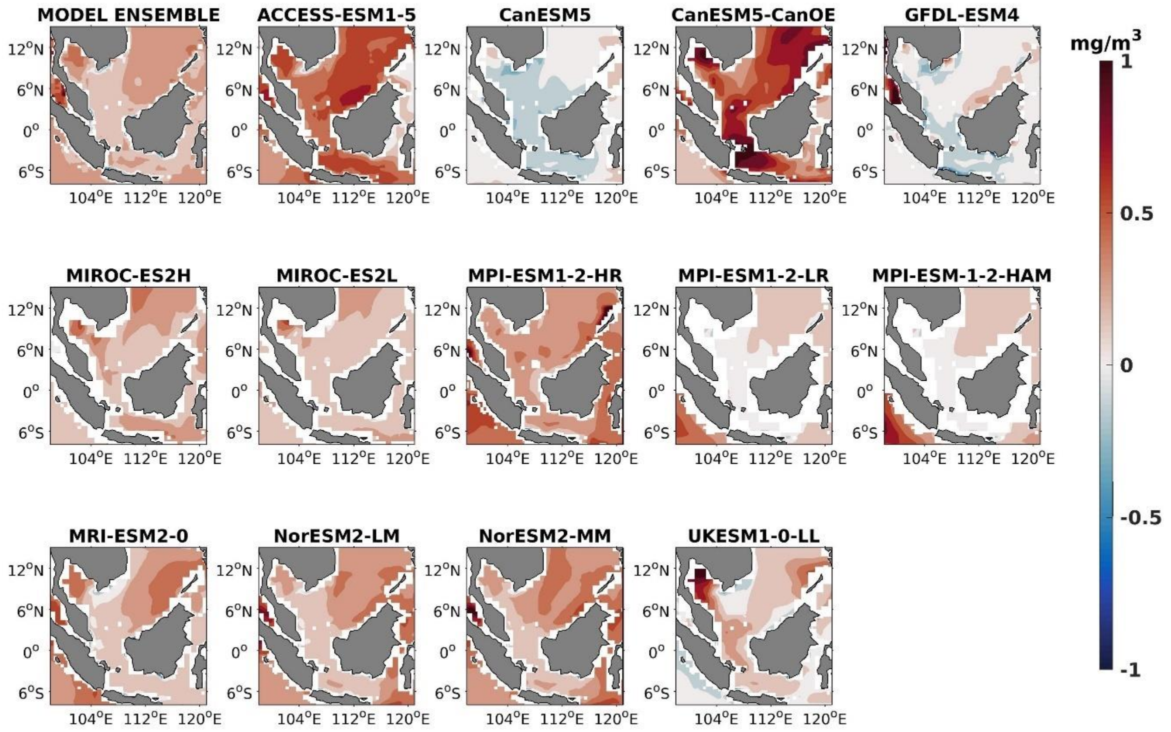

Figure 1. DJF spatial biases of surface chlorophyll for 13 individual models and model ensemble relative to reference.

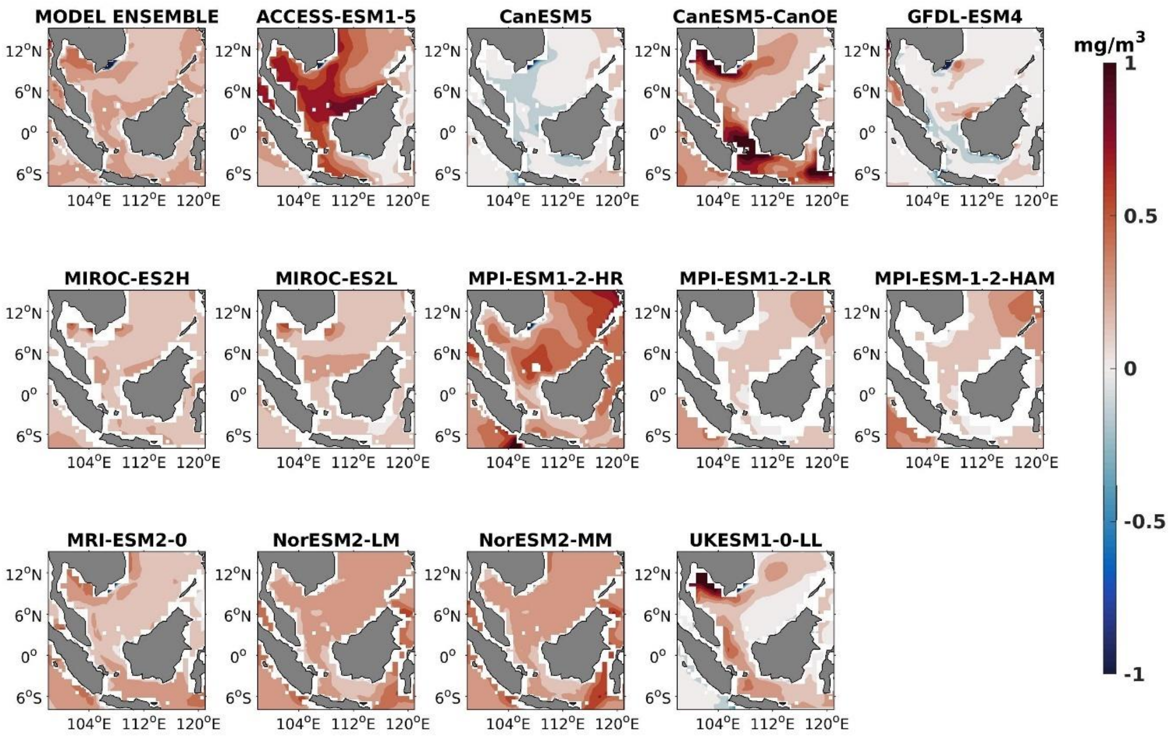

Figure 2. Same as Fig.2 but for JJA.

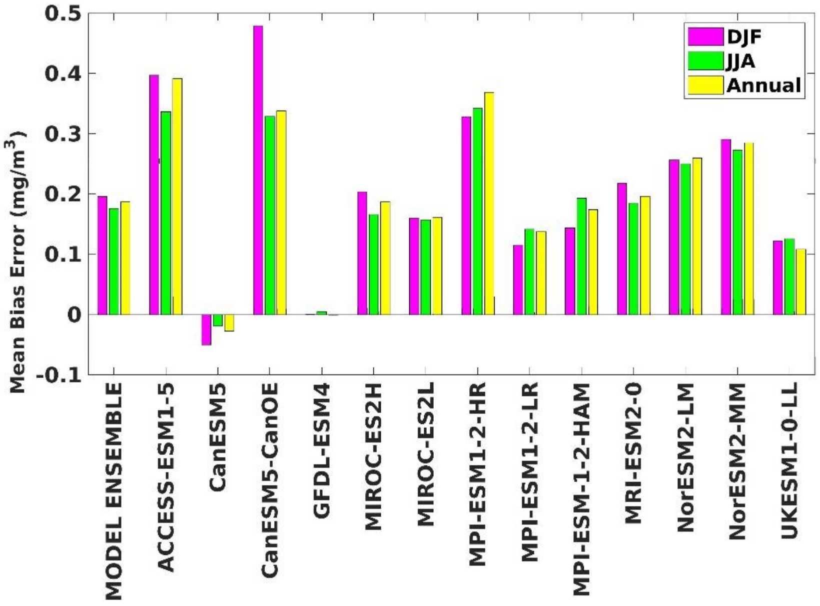

Figure 3. The mean bias of surface chlorophyll for both seasons (DJF, JJA) and annual.

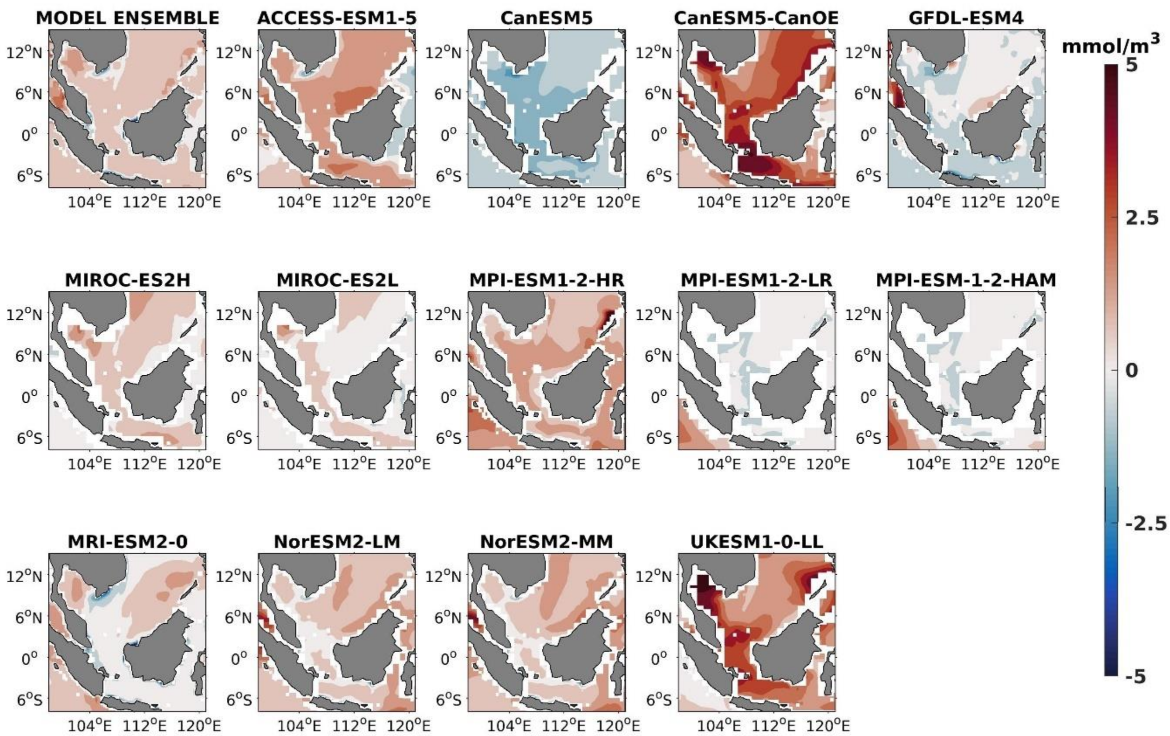

Figure 4. DJF spatial biases of surface phytoplankton for 13 individual models and model ensemble relative to reference.

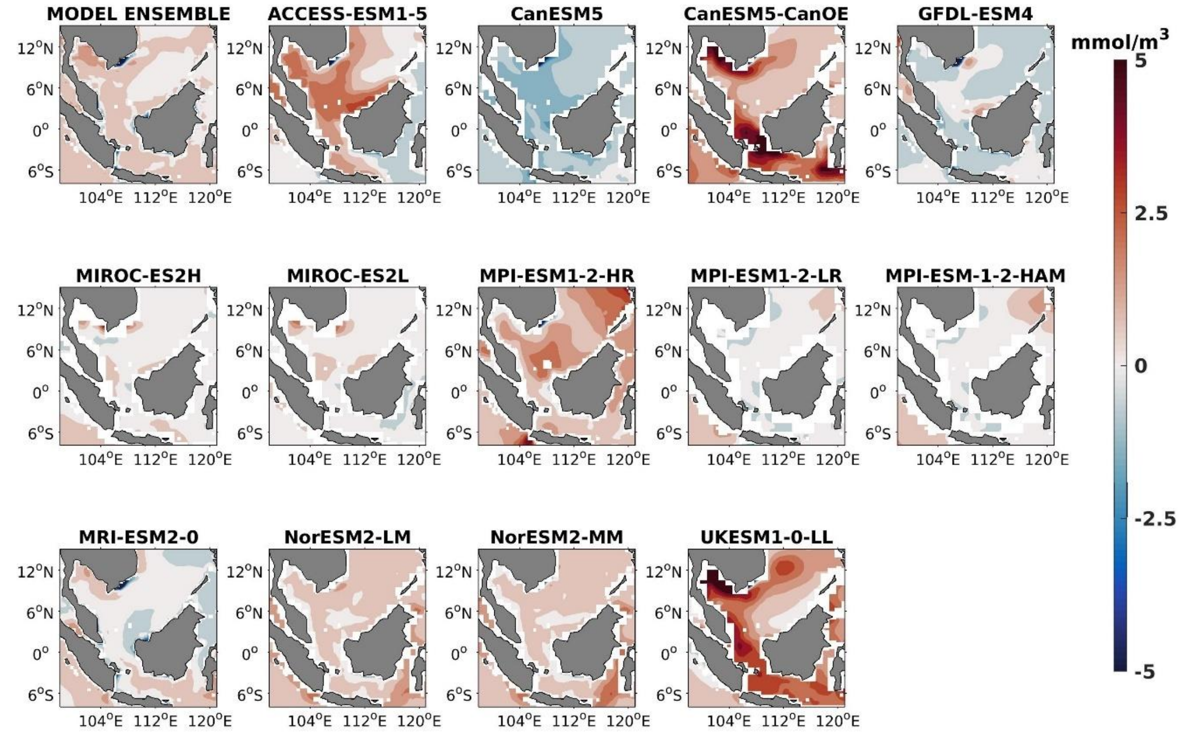

Figure 5. Same as Fig. 5 but for JJA.

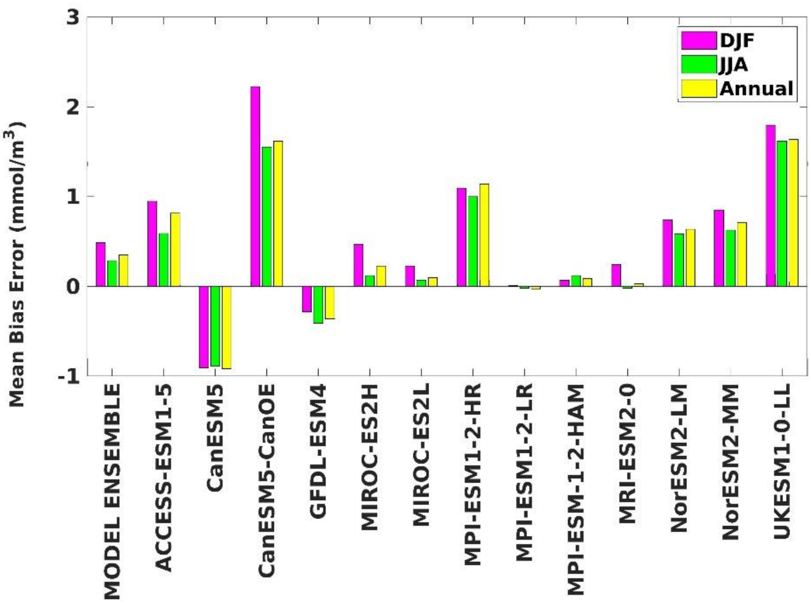

Figure 6. The mean bias of surface phytoplankton for both seasons (DJF, JJA) and annual.

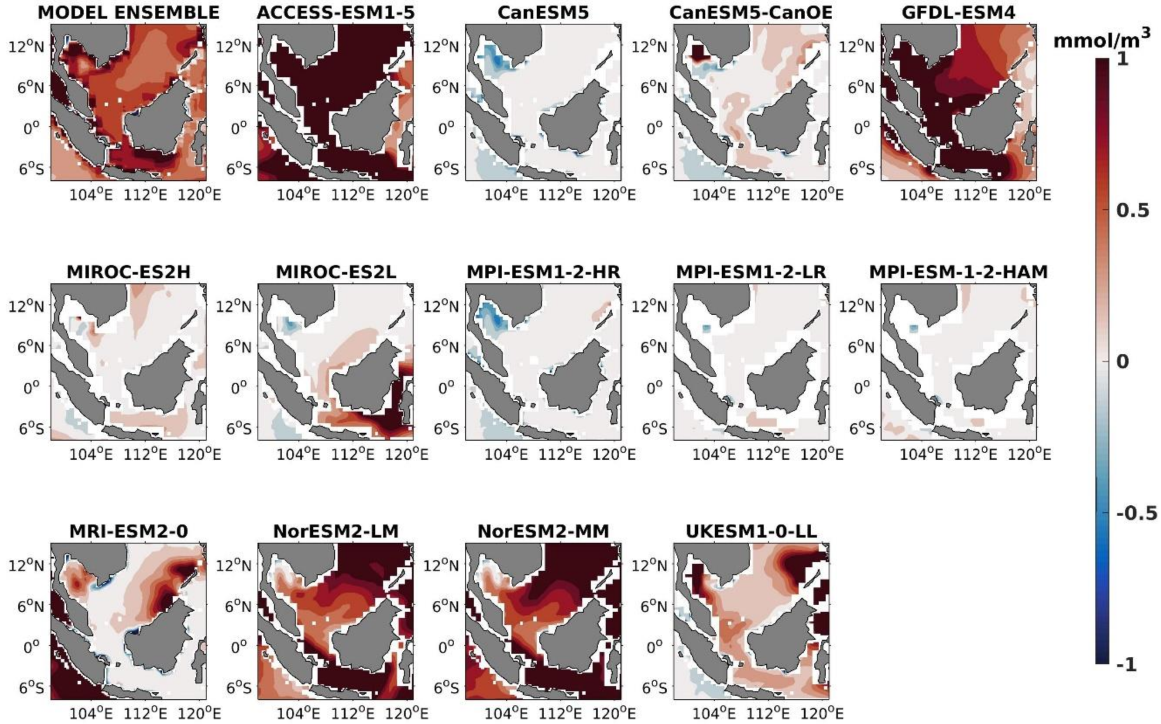

Figure 7. DJF spatial biases of surface nitrate for 13 individual models and model ensemble relative to reference.

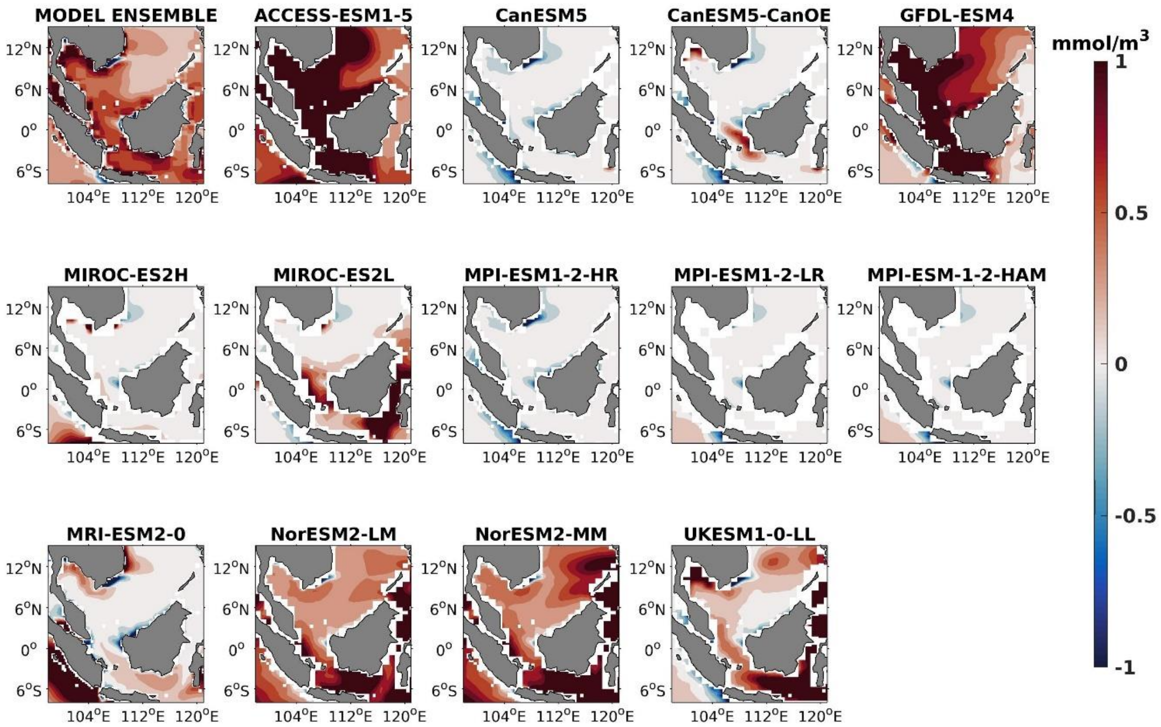

Figure 8. Same as Fig. 8 but for JJA.

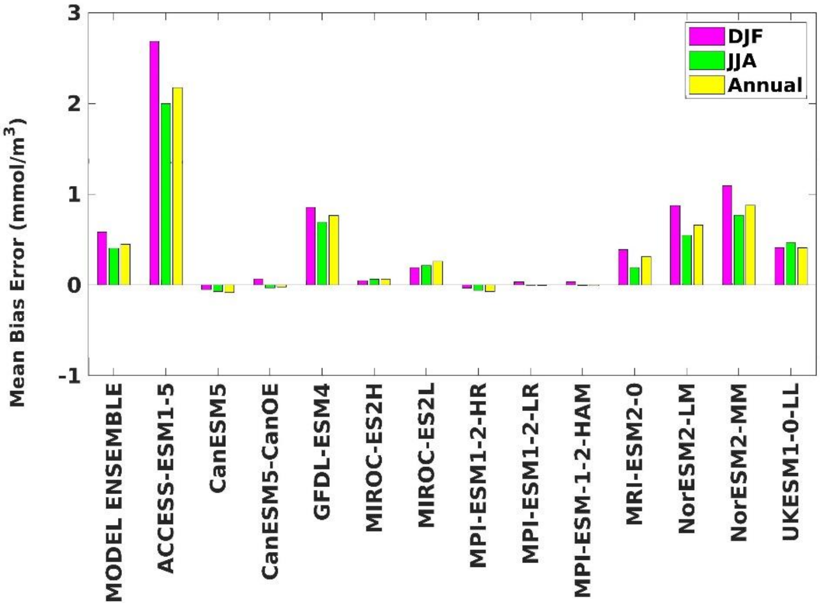

Figure 9. The mean bias of surface nitrate for both seasons (DJF, JJA) and annual.

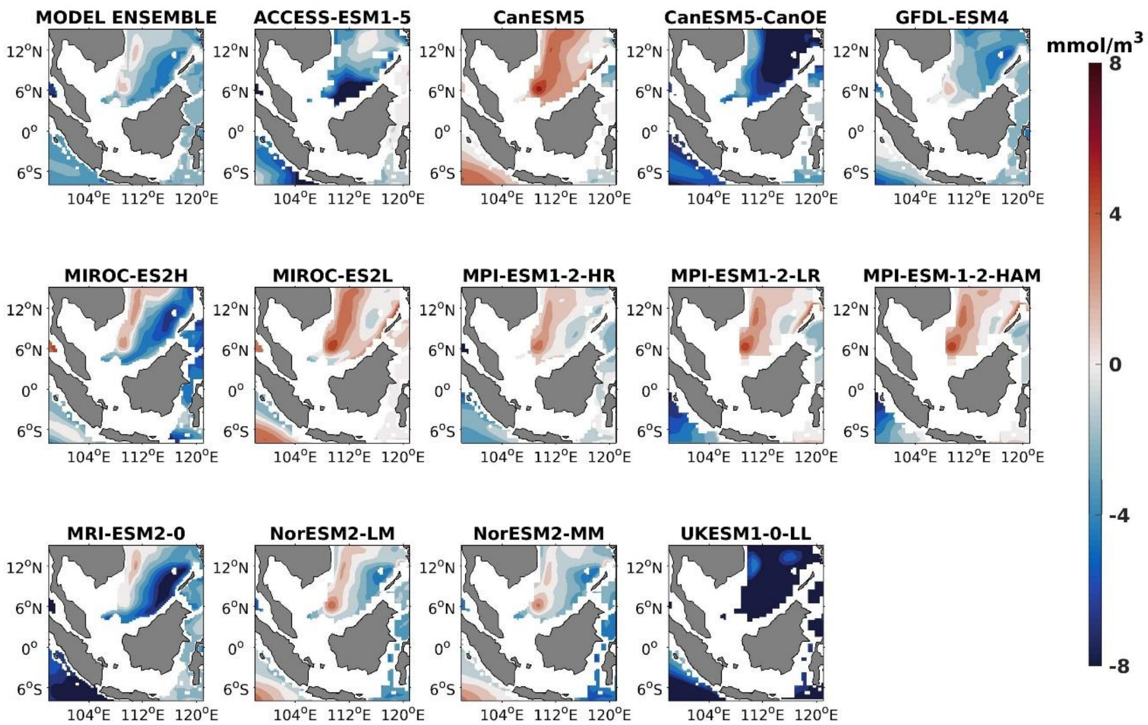

Figure 10. DJF spatial biases of nitrate at 70-meter depth for 13 individual models and model ensemble relative to reference.

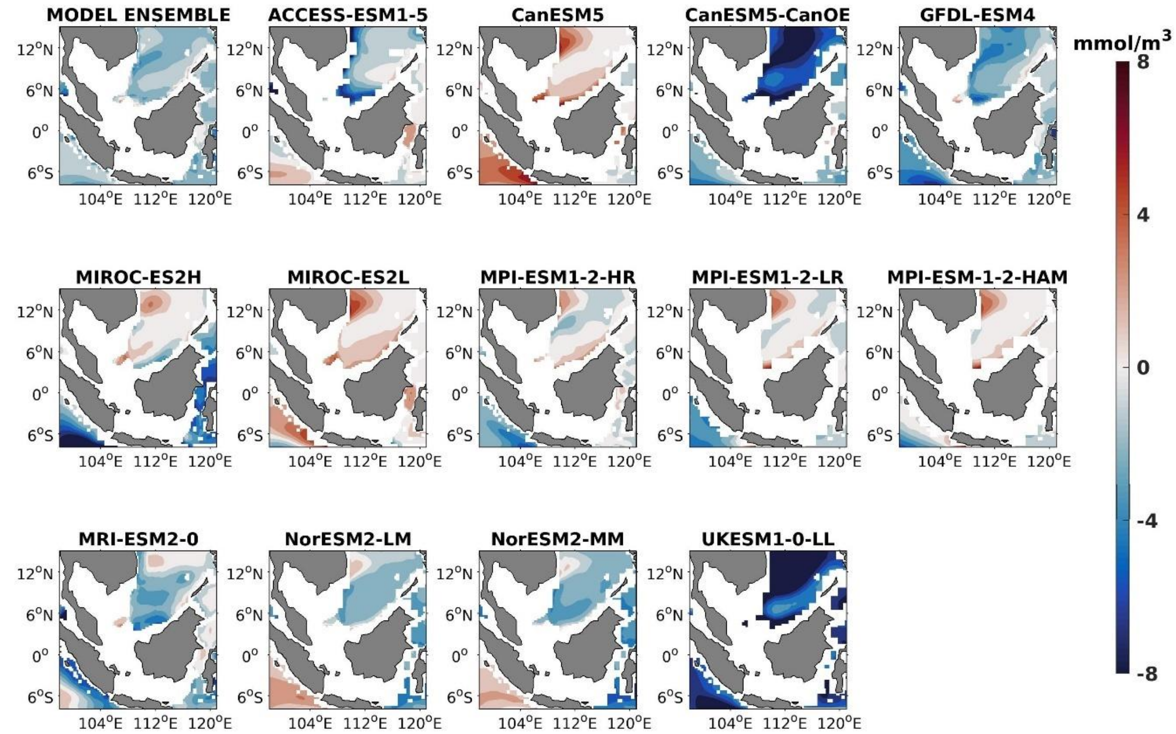

Figure 11. Same as Fig. 11 but for JJA.

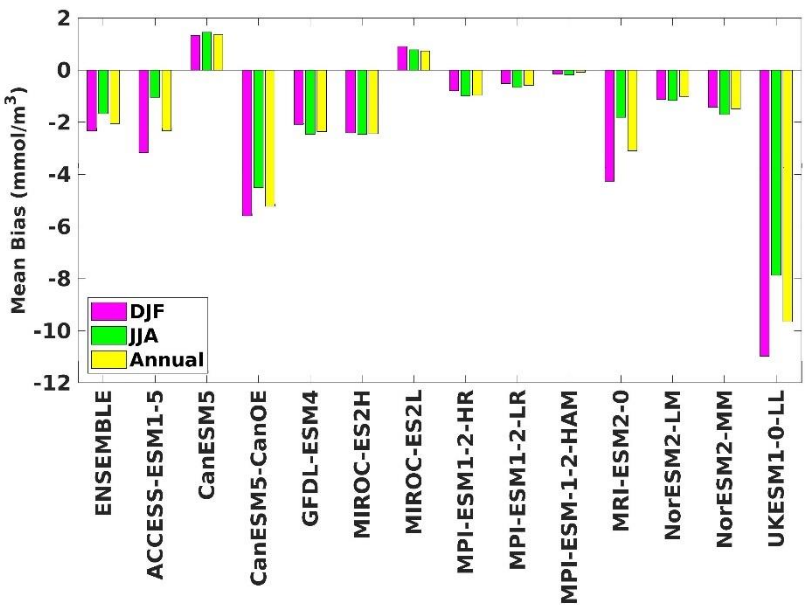

Figure 12. The mean bias of nitrate at 70-meter depth for both seasons (DJF, JJA) and annual.

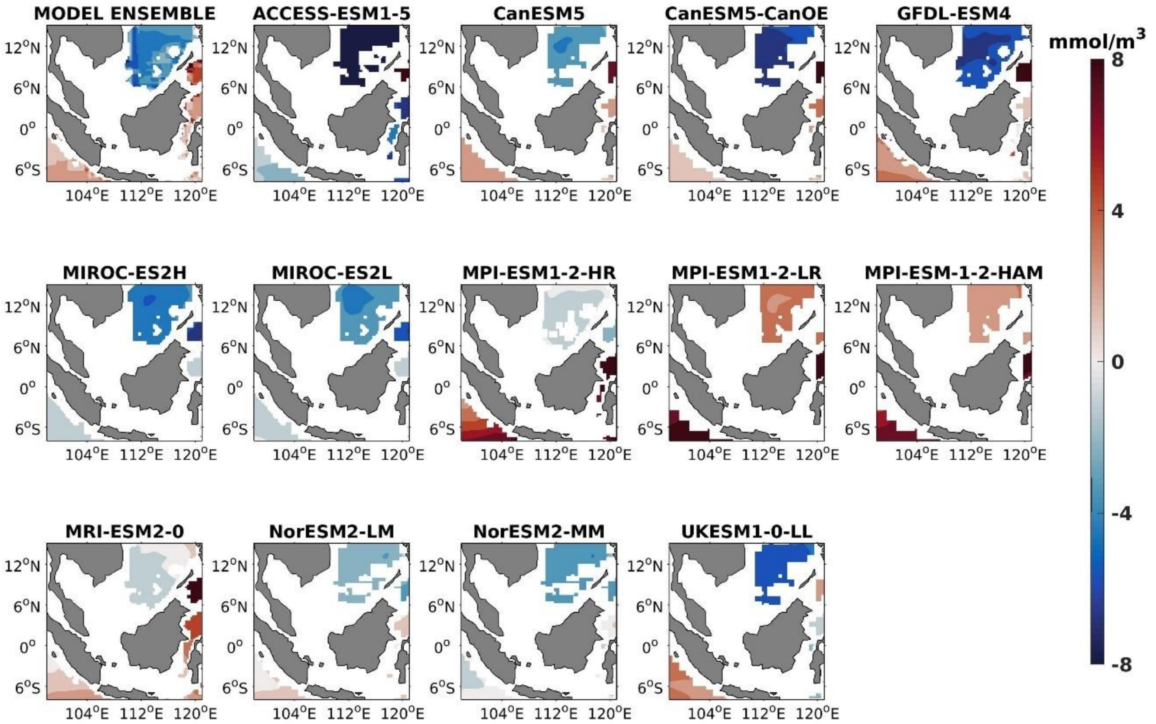

Figure 13. DJF spatial biases of nitrate at 1000-meter depth for 13 individual models and model ensemble relative to reference.

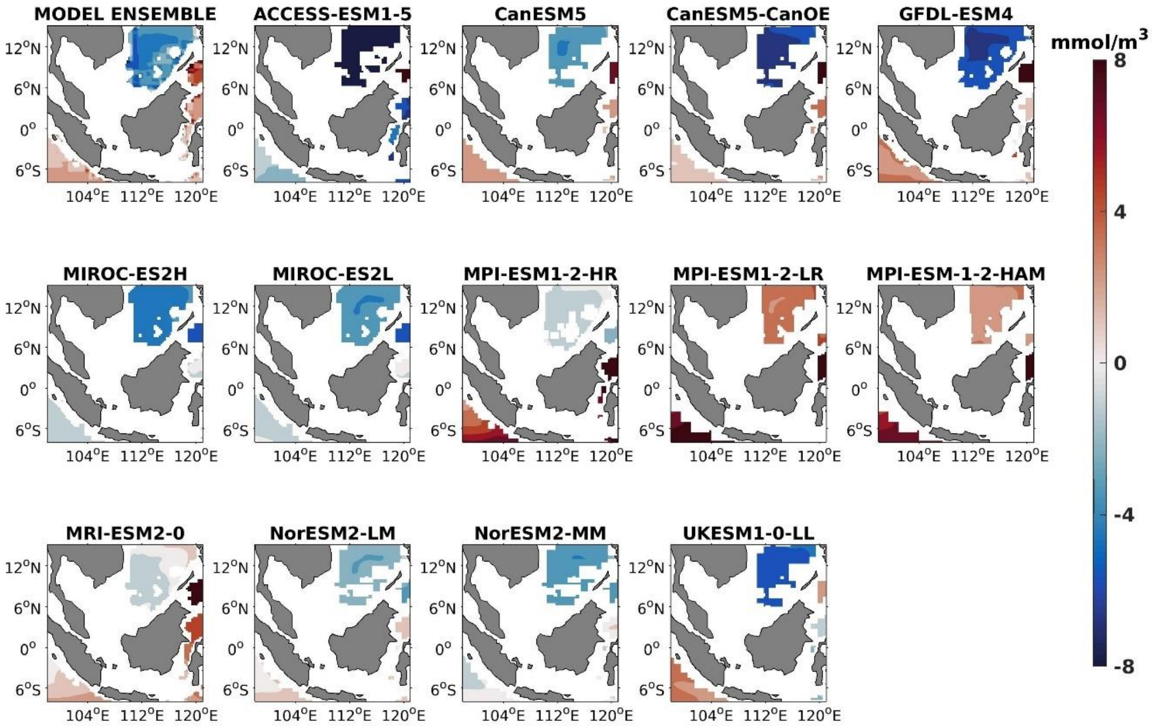

Figure 14. Same as Fig. 14 but for JJA.

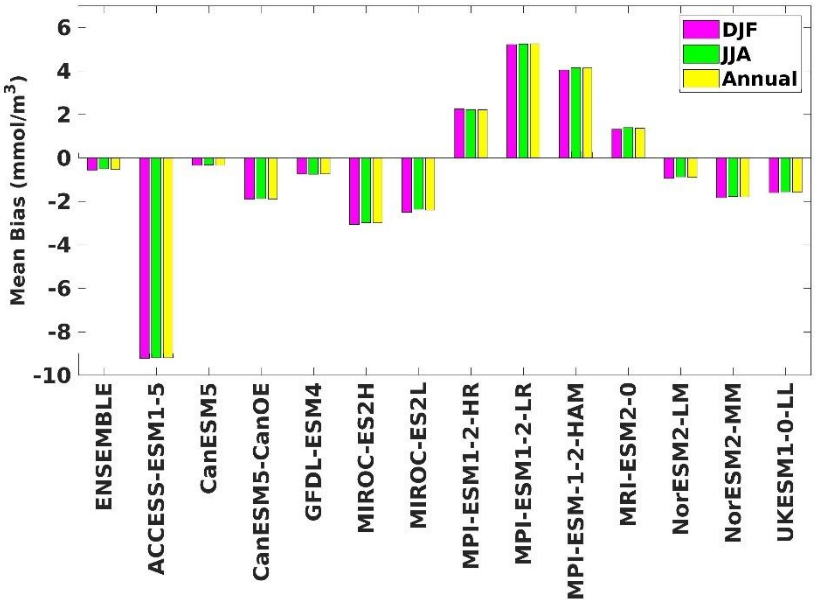

Figure 15. The mean bias of nitrate at 1000-meter depth for both seasons (DJF, JJA) and annual.

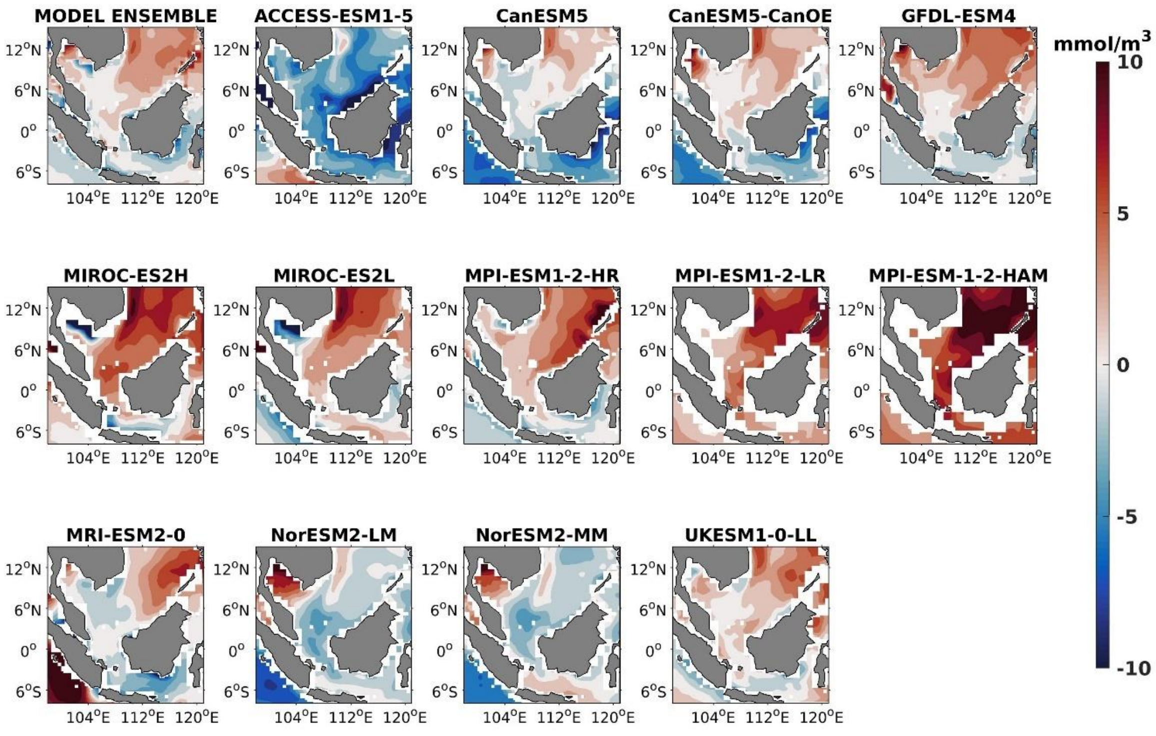

Figure 16. DJF spatial biases of surface oxygen for 13 individual models and model ensemble relative to reference.

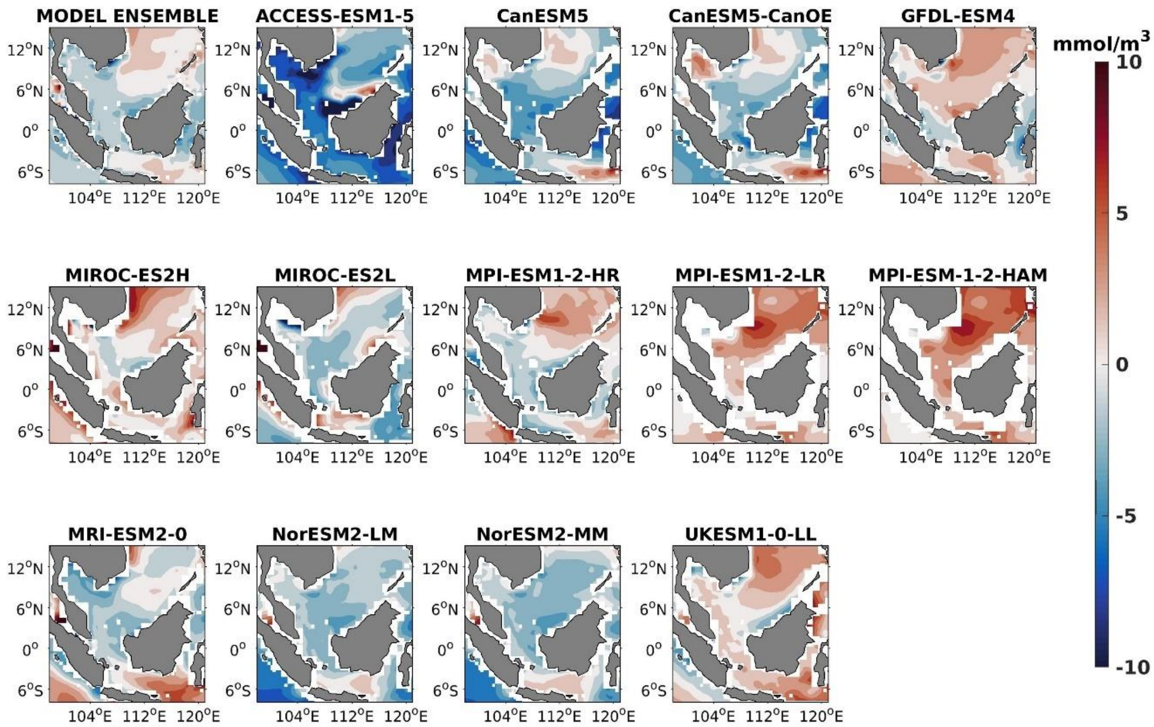

Figure 17. Same as Fig. 17 but for JJA.

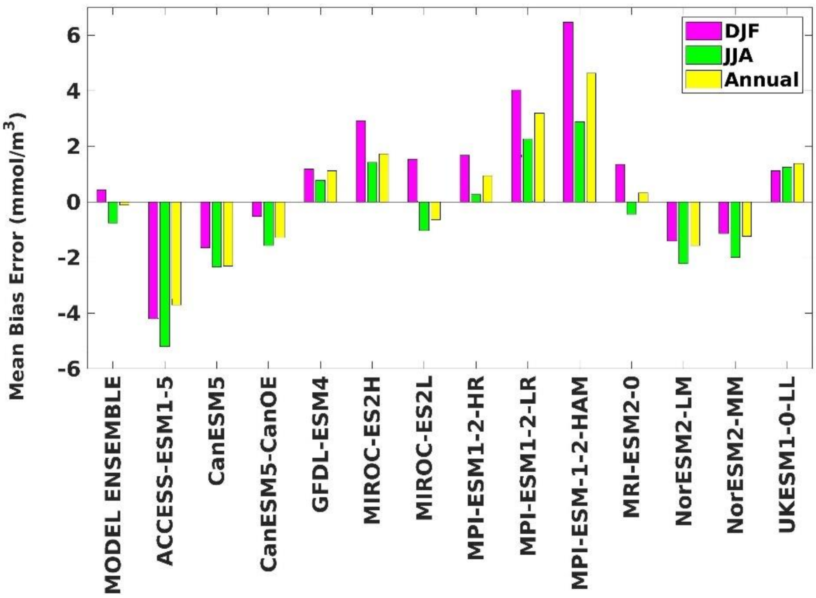

Figure 18. The mean bias of surface oxygen for both seasons (DJF, JJA) and annual.

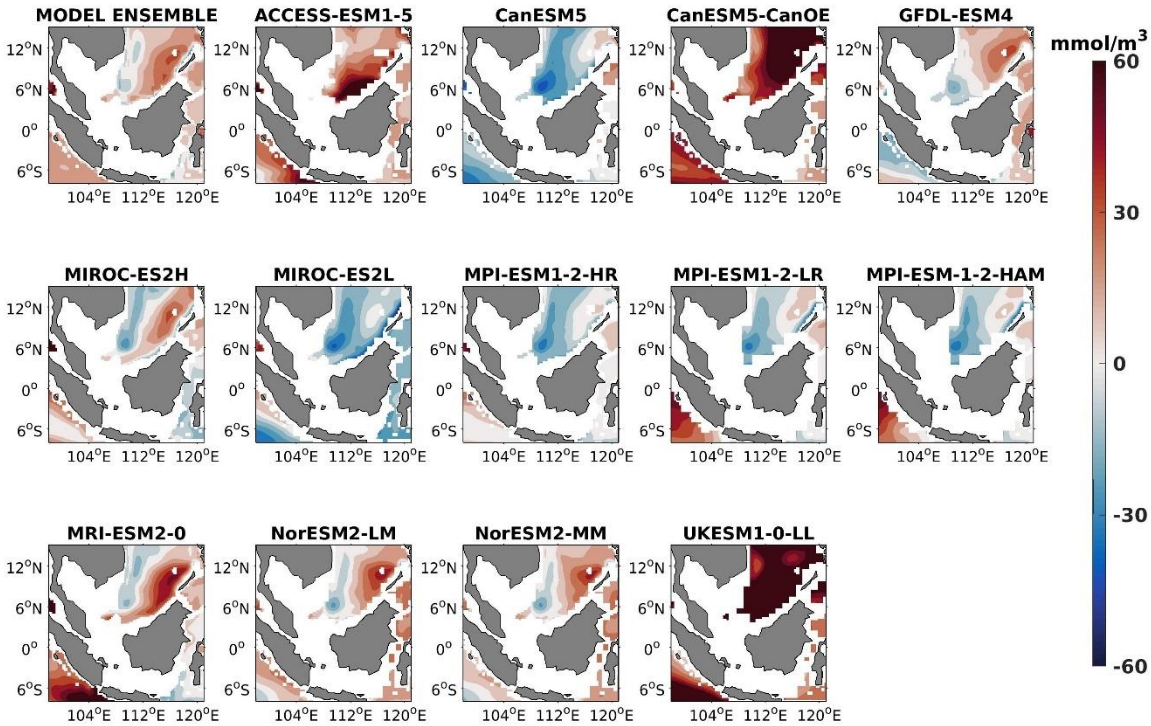

Figure 19. DJF spatial biases of oxygen at 70-meter depth for 13 individual models and model ensemble relative to reference.

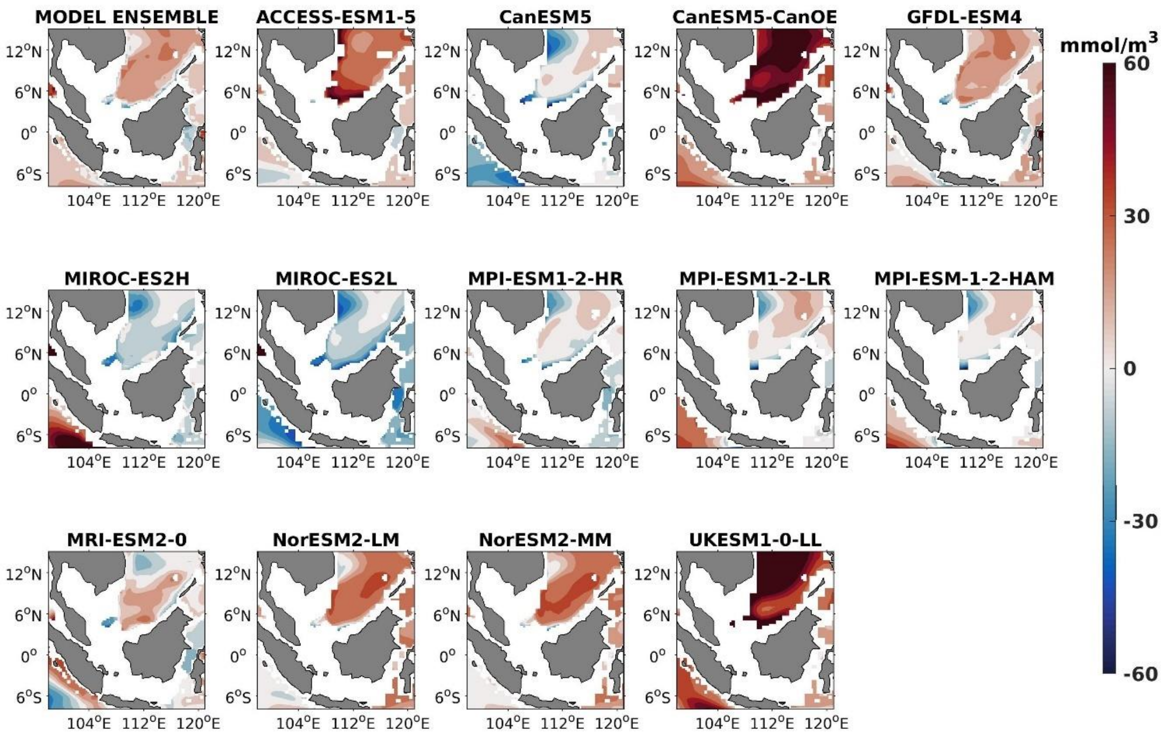

Figure 20. Same as Fig. 20 but for JJA.

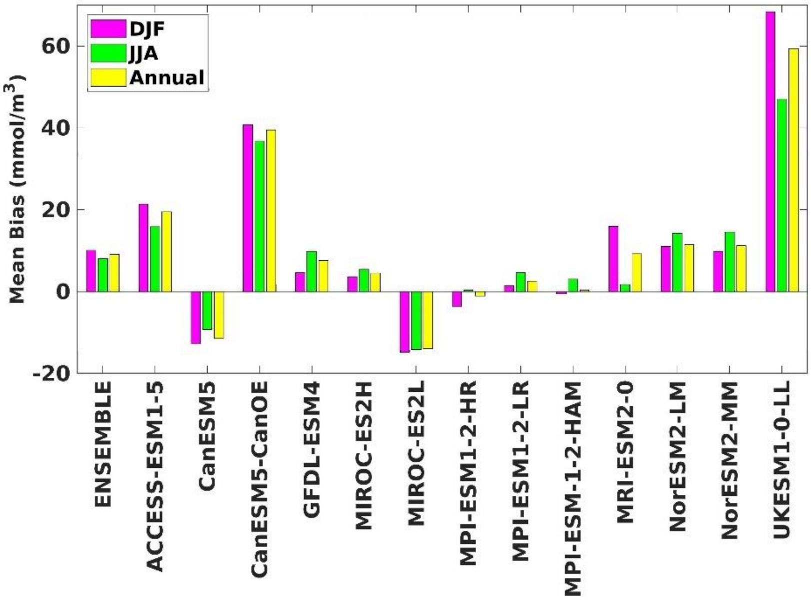

Figure 21. The mean bias of oxygen at 70-meter depth for both seasons (DJF, JJA) and annual.

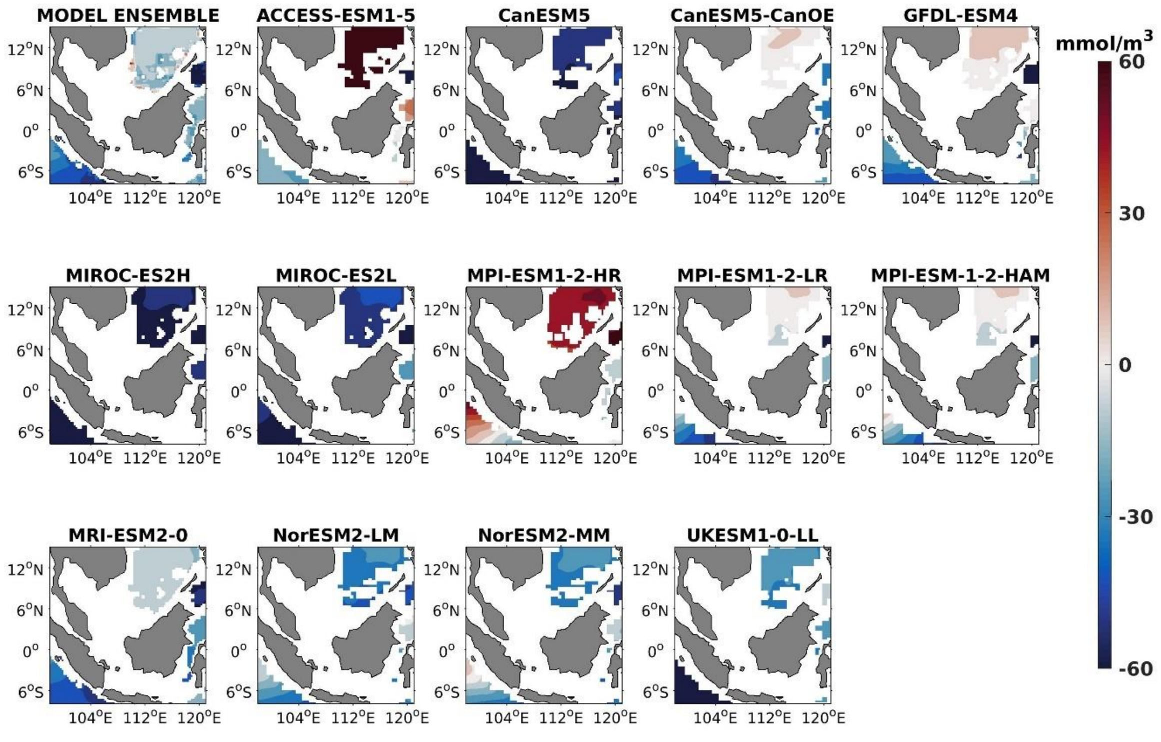

Figure 22. DJF spatial biases of oxygen at 1000-meter depth for 13 individual models and model ensemble relative to reference.

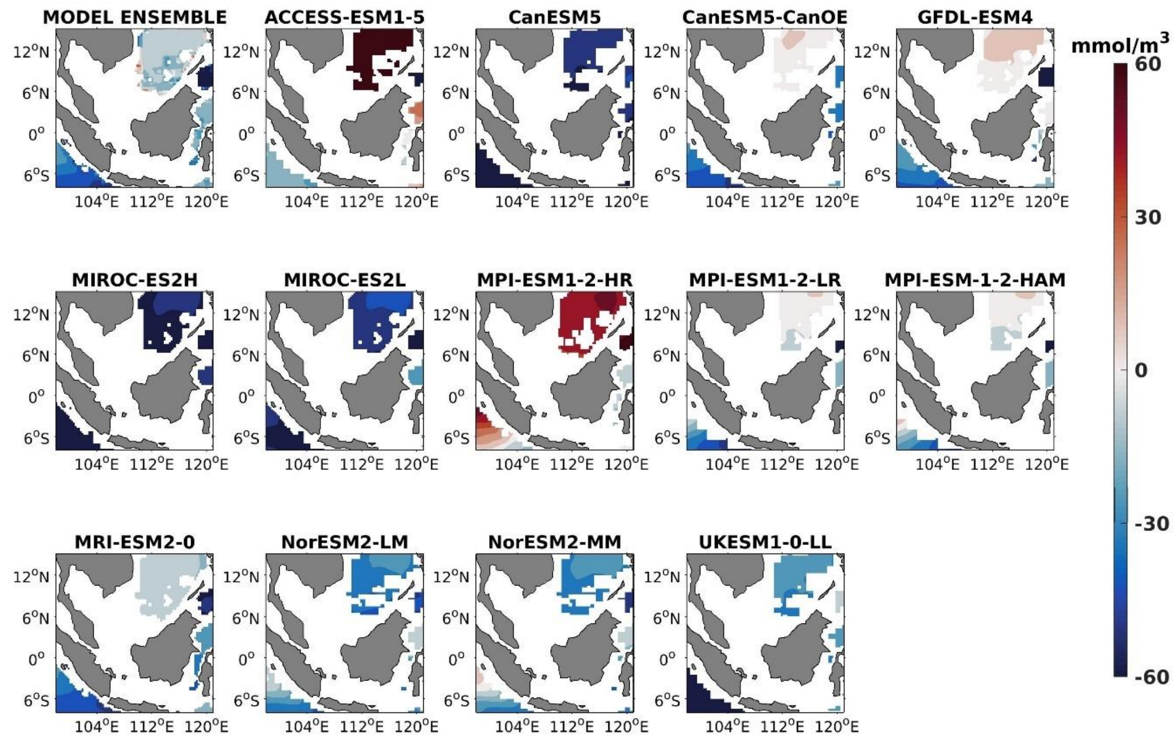

Figure 23. Same as Fig. 23 but for JJA.

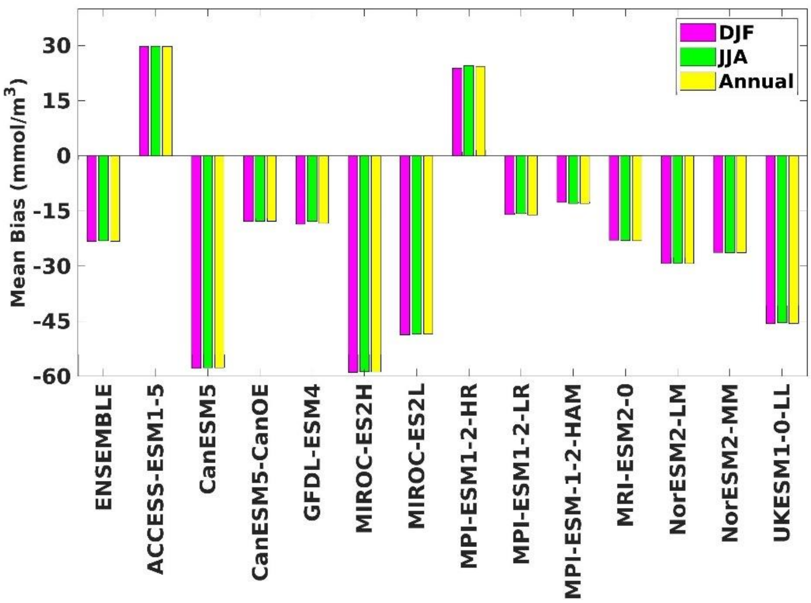

Figure 24. The mean bias of oxygen at 1000-meter depth for both seasons (DJF, JJA) and annual.

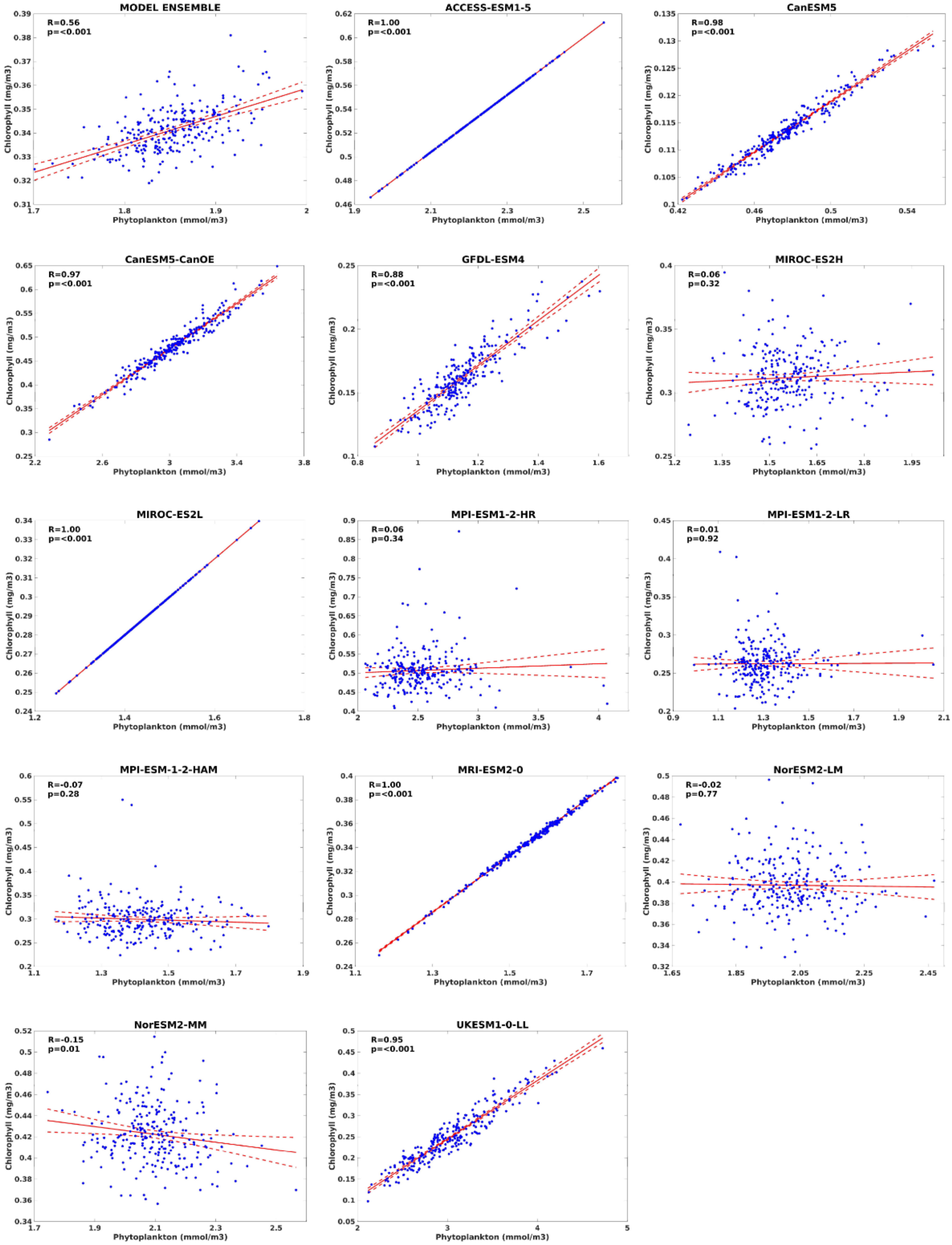

Figure 25. Relationships between chlorophyll and phytoplankton for model ensemble and 13 individual models in southern South China Sea during the study period (1993 – 2014). Dashed red lines represent 95% confidence interval with 0.05 as the level of significance (α).

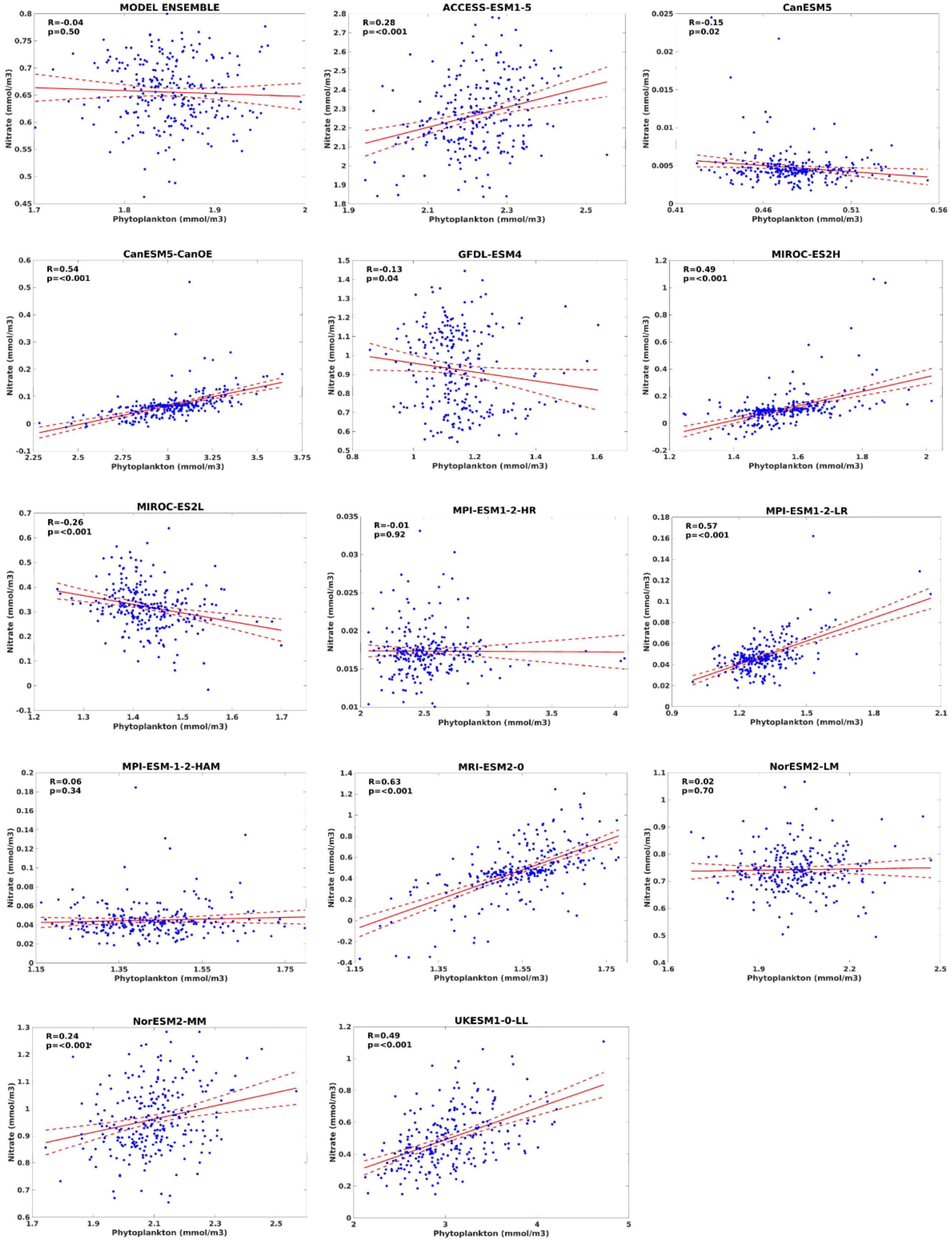

Figure 26. Relationships between nitrate and phytoplankton for model ensemble and 13 individual models in southern South China Sea during the study period (1993 – 2014). Dashed red lines represent 95% confidence interval with 0.05 as the level of significance (α).

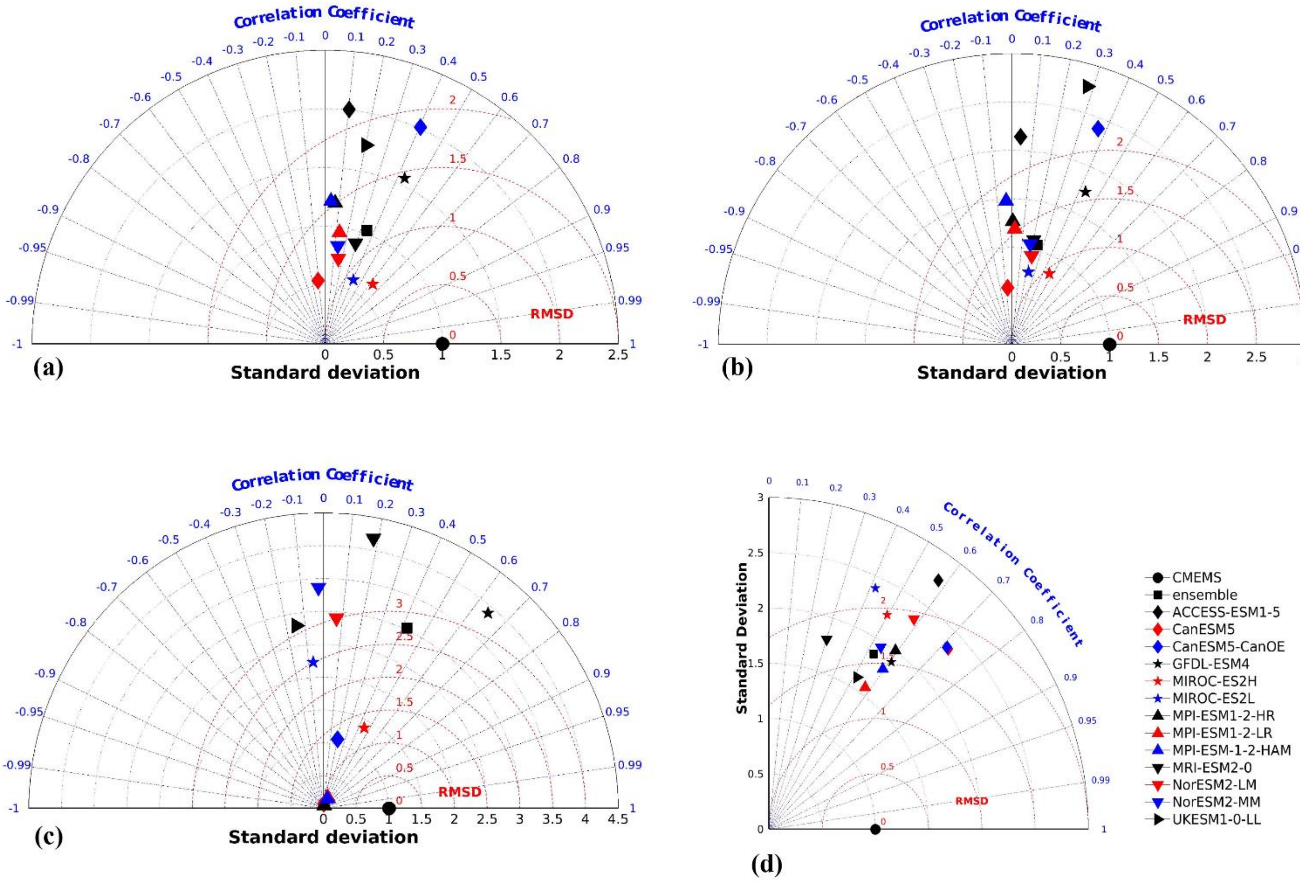

Figure 27. Annual Taylor Diagram for (a) chlorophyll, (b) phytoplankton, (c) nitrate and (d) oxygen.

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