

General comments:

The paper has been improved by incorporating a more in-depth discussion on the model structure, resolution, and physical processes. However, I still have some concerns regarding the datasets used in the model. If this study aims to conduct an evaluation, it should consider actual observations such as satellite products.

I understand that CMEMS datasets have been evaluated and shown good agreement with the World Ocean Atlas and ocean colour data. Nevertheless, I am curious as to why the authors did not use the available observations that was mentioned in the author's response (i.e. in-situ measurements taken from the Southern South China Sea). If these datasets are not available for the entire time series (1993-2014), could the authors consider using the specific years where observations are available?

Additionally, the manuscript could benefit from further clarity. Some sections are challenging to follow due to long paragraphs that are lacking a specific point, especially some parts of the introduction as well as the results and discussion section (chlorophyll and phytoplankton carbon, nitrate, oxygen, and the entire section 4.2). I recommend having someone outside the research group to review the manuscript for readability.

Due to the absence of observational data, I am currently unable to recommend this manuscript for publication.

Specific comments:

Line 17: The authors mostly refer southwest monsoon as JJA and northeast monsoon as DJF, so this is not representative of the text, please change to June-August and December-February.

Line 133-134: Can you provide a reference for this sentence

Line 148: I disagree, there are satellite products which count as an observation for chlorophyll (even primary production).

Line 153: Wahyudi et al., 2023 is missing from the reference list

Line 155: Triana et al., 2021 is missing from the reference list

Line 157: Chen et al., 2023 is missing from the reference list

In figures with maps, the authors show the model ensemble bias, but there are no specific comments about it in the text. Some of the figures and subfigures are also not referenced in the text.

In line 227-229, the authors mentioned that errors larger than $+0.1 \text{ mg/m}^3$ indicates a notable discrepancy, but in line 236-238, the authors said that $>0.15 \text{ mg/m}^3$ is within acceptable range – so the three models mentioned earlier are still within acceptable range?

Line 241-242: Do you mean: does not overestimate?

Line 253-268: great discussion!

Line 308-309: Do you mean: ranging from?

Perhaps the authors can show the ESM's nitrate profiles to show whether ESM can capture where the nitracline is during different season? (is this the dissolved inorganic nitrogen?)

Line 338: is this DJF?

Line 348-350: Can you provide reference for this?

Line 434: you mean all the biogeochemistry variables? Because some models can reproduce the pattern of at least one of the variables.

Section 4.2.2 – can you also make some comments on using a better phytoplankton parameterisation such as the nutrient quota? (or flexible N:C ratio of phytoplankton).

Line 520 – Taylor's diagram or Taylor diagram? Please be consistent.

Line 580-581 – You are repeating line 579 – 580.

Line 618 - Perhaps i am missing something but the authors have not mentioned about annual scales at all in the first half of the results and discussion; and is only touched in the Taylor diagram part

Line 620 – is this only at the surface? Perhaps also consider the deeper depths as well because this is the conclusion section.

Line 634 – do not use *etc.*