The paper by Marshal et al. evaluates how well 13 CMIP6 ESMs simulate biogeochemical variables in the southern South China Sea (SCS) by comparing their simulation of chlorophyll, phytoplankton biomass, nitrate, and oxygen with CMEMS data. The authors use statistical metrics, such as correlation coefficient, RMSD, and mean bias error, to evaluate and rank which model can represent CMEMS data at southern SCS the best. They found that although most ESMs capture the observed seasonal pattern of each biogeochemical property, some models exhibit overestimation or underestimation of their magnitude. The paper provides an important evaluation of the CMIP6 ESMs in the southern SCS, with a thorough assessment of phytoplankton, nutrient, and oxygen -related biogeochemical properties. However, I have some issues with the choice of the reference dataset and the structure of the manuscript.

General comments:

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. 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.

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.

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.

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).

Specific comments:

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

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

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)

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

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

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

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

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

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)?

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

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

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

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

L184, 219 – what is acceptable range?

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

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

L251 – Can you give example of the important processes?

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?

Figure 6 could do with higher resolution.

References:

Racault, M.-F., Raitsos, D.E., Berumen, M.L., Brewin, R.J.W., Platt, T., Sathyendranath, S., Hoteit, I., 2015. Phytoplankton phenology indices in coral reef ecosystems: Application to ocean-color observations in the Red Sea. Remote Sensing of Environment 160, 222–234. https://doi.org/10.1016/j.rse.2015.01.019

Gittings, J.A., Raitsos, D.E., Kheireddine, M. *et al.* Evaluating tropical phytoplankton phenology metrics using contemporary tools. *Sci Rep* **9**, 674 (2019). https://doi.org/10.1038/s41598-018-37370-4

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

Sinha, B., Buitenhuis, E.T., Quéré, C.L., Anderson, T.R., 2010. Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models. Progress in Oceanography 84, 204–224. https://doi.org/10.1016/j.pocean.2009.10.003

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J., Lenton, A., Li, H., Lovenduski, N., Orr, J., Palmieri, J., Schwinger, J., Séférian, R., Stock, C., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., Ziehn, T., 2020. Twenty-first century ocean warming, acidification, deoxygenation, and upper ocean nutrient decline from CMIP6 model projections. Biogeosciences Discussions 1–43. https://doi.org/10.5194/bg-2020-16

Hinrichs, C., Köhler, P., Völker, C., and Hauck, J.: Alkalinity biases in CMIP6 Earth system models and implications for simulated CO_2 drawdown via artificial alkalinity enhancement, Biogeosciences, 20, 3717–3735, https://doi.org/10.5194/bg-20-3717-2023, 2023.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. 8, 2465–2513 (2015).

Lovato, T., Peano, D., Butenschön, M.,Materia, S., Iovino, D., Scoccimarro, E.,et al. (2022). CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2). Journal of Advances in Modeling Earth Systems, 14, e2021MS002814. https://doi.org/10.1029/2021MS002814

Long, M. C. et al. Simulations with the marine biogeochemistry library (MARBL). J. Adv. Model. Earth Syst. 13, e2021MS002647 (2021).