We sincerely thank the Anonymous Referee #2 for the thorough and constructive comments on our manuscript. We appreciate the effort taken to highlight both the strengths and the shortcomings of our study, particularly regarding the use of observational data sets and the robustness of the model ranking exercise. Based on the instructions of the editorial support team of Copernicus Publications, we will provide a point-by-point response and outline the revisions we will undertake below before we make revision on the manuscript. Italic font will be used to distinguish our replies from the reviewer' comments.

The manuscript "Evaluating the performance of CMIP6 models in simulating Southern Ocean biogeochemistry" analyzes coupled carbon-climate Earth system model fidelity for surface chlorophyll, nitrate, silicate, and iron, the deep chlorophyll maximum, and particular organic carbon across subregions of the Southern Ocean to rank the models which is a highly valuable analysis given the historical challenges in both observations collection and model fidelity and importance of the Southern Ocean for heat and carbon uptake. The biggest weakness of the current manuscript is the assumption that inter-model differences and biases should be attributed to biogeochemical formulation rather than the underlying physics, including representation of temperature, mixed layer depth, upwelling, and upper ocean stratification, transport, and turbulence, all of which are long standing challenges in the climate community. While a detailed discussion of the potential of physical biases and their potential implications is outside the scope of the present manuscript, the possibility of physical attribution should be mentioned. Otherwise I have only minor comments.

We will mention that the underlying physics of each will be mentioned and that this can influence the performance of the BGC component of the model, especially in coastal system .

Specific comments by line number:

9 - This assertion is highly overstated - see lines 67-82 which contradict this as well as such literature as:

Frölicher, T.L., Sarmiento, J.L., Paynter, D.J., Dunne, J.P., Krasting, J.P. and Winton, M., 2015. Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. Journal of Climate, 28(2), pp.862-886.

Mongwe, N.P., Vichi, M. and Monteiro, P.M., 2018. The seasonal cycle of p CO 2 and CO 2 fluxes in the Southern Ocean: diagnosing anomalies in CMIP5 Earth system models. Biogeosciences, 15(9), pp.2851-2872.

Rickard, G.J., Behrens, E., Chiswell, S., Law, C.S. and Pinkerton, M.H., 2023. Biogeochemical and physical assessment of CMIP5 and CMIP6 ocean components for the southwest Pacific Ocean. Journal of Geophysical Research: Biogeosciences, 128(5), p.e2022JG007123.

Nevison, C.D., Manizza, M., Keeling, R.F., Stephens, B.B., Bent, J.D., Dunne, J., Ilyina, T., Long, M., Resplandy, L., Tjiputra, J. and Yukimoto, S., 2016. Evaluating CMIP5 ocean biogeochemistry and Southern Ocean carbon uptake using atmospheric potential oxygen: Present-day performance and future projection. Geophysical Research Letters, 43(5), pp.2077-2085.

We will change open sentence of the abstract "yet the quality of its representation in Earth System Models (ESMs) remains unquantified" to "yet comprehensive assessments of its representation in Earth System Models (ESMs) are still limited"

48 - which? High iron requirement?

More favourable nutrient condition is iron and silicon supply.

60 - by "integration of". do the authors mean "assessment with"? It is not clear what "data" is integrated into these models to represent the Southern Ocean except for topography and radiative forcing.

Models are constrained by some observed dataset. We will change "integration of" to "constrained by".

99 - Why define the acronym when it is not used again until the acknowledgments and also defined there?

We will remove "(NCI)" at line 99 and line 563.

Eq 1-4 - These are all pretty common statistical definitions which could be removed for space.

We will remove these equations.

167 – "MPI-ESM models" should be "MPI-ESMs"

We will change "MPI-ESM models" to "MPI-ESMs" at line 167, 250, 300, 337, and 386.

225 - Should be "Fig. 1" to point to chlorophyll.

We will change "Fig. 5" at line 225 to "Fig. 1".

510-558 - The attribution here to biological complexity seems to assume that the Southern Ocean physics that drives the biogeochemistry is perfect in these models. This is not the case and is the subject of many papers. Much of the focus has been on wind and sea ice biases and upper ocean stratification (e.g. Beadling et al, 2020), temperature, (Luo et al., 2023 and polynya (Mohrmann et al., 2021):

Beadling, R. L., Russell, J. L., Stouffer, R. J., Mazloff, M., Talley, L. D., Goodman, P. J., ... & Pandde, A. (2020). Representation of Southern Ocean properties across coupled model intercomparison project generations: CMIP3 to CMIP6. Journal of Climate, 33(15), 6555-6581.

Luo, F., Ying, J., Liu, T., & Chen, D. (2023). Origins of Southern Ocean warm sea surface temperature bias in CMIP6 models. npj Climate and Atmospheric Science, 6(1), 127.

Mohrmann, M., Heuzé, C., & Swart, S. (2021). Southern Ocean polynyas in CMIP6 models. The Cryosphere, 15(9), 4281-4313.

We accept the reviewer's critique that this section over-attributes biases to biological complexity without considering physical drivers. Although a detailed discussion of the potential of physical biases and their potential implications is outside the scope of the manuscript, we will add some content to discuss the effect of physical attributes on biogeochemical performance.