We thank the Reviewer for their positive feedback and valuable inputs on the manuscript. Below, we provide a detailed response to each of the comments — reviewer comments are in black, and our responses are in blue.

**General comments:**

The writing is clear and the manuscript introduces new findings that contribute to our understanding of chlorophyll-a dynamics. I recommend accepting the manuscript with minor revisions. The following comments are merely provided as suggestions to further improve the manuscript's completeness and clarity.

We thank the reviewer for their encouraging assessment of our manuscript.

**Summary:**

This study compares simulated surface chlorophyll-a (Schla) variability from a subset of CMIP6 Earth System Models (ESMs) with satellite observations and contrasts this performance with that of SST. The analyses highlight discrepancies in the ability of ESMs to simulate Schla across different timescales, with a specific focus on the understudied sub-seasonal variability. The ESM simulations are selected based on the availability of daily Schla and SST outputs. Temporal variability is decomposed into sub-seasonal, seasonal, and multi-annual scales, identifying three main groups: one showing an overestimation of sub-seasonal variability which is attributed to the coarse spatial resolution of the ESMs, a second group showing an underestimation of sub-seasonal variability, potentially linked to intrinsic predator-prey oscillations within the ESMs, and a third group displaying an overestimation of total variance but consistent temporal decomposition. The authors conclude that, unlike SST, ESMs do not adequately represent Schla variability, emphasizing the need for additional CMIP simulations with higher spatial and temporal resolutions to address these limitations.

**Specific comments:**

Perhaps the manuscript could include an explicit mention of the limitations of the approaches and how they affect the final findings, specifically concerning:

- The biases/uncertainties of satellite observations: While the manuscript uses satellite observations as the benchmark for comparison, it would strengthen the discussion to acknowledge the inherent biases and limitations of these datasets. For instance, biases introduced by gap-filling and uncertainties in the retrieval process could affect the representation of SChla variability. It would be beneficial if the authors discussed these biases and how they might influence the overall findings.

We will add a new paragraph highlighting the biases and uncertainties in satellite observations in line 275: "It should be noted that biases introduced by gap-filling in satellite-derived data can lead to an inaccurate representation of SChl variability, as missing or interpolated data points may not capture the true temporal or spatial patterns of chlorophyll concentrations. Additionally, uncertainties in the retrieval process, such as atmospheric corrections and sensor calibration, can further distort the observed variability, affecting the reliability of satellite-derived estimates of surface chlorophyll. However, satellite ocean color measurements remain the only available source of high-frequency observations of SChl over extended periods at a global scale. Furthermore, a comparison of SChl at a mooring location in the BOUSSOLE in

the Gulf of Lion showed that satellites can capture SChl variability at higher temporal resolutions reasonably well (Keerthi et al., 2020). Nonetheless, cloud cover remains a limitation that can affect the accuracy of these measurements."

Keerthi, M. G., Levy, M., Aumont, O., Lengaigne, M. & Antoine, D.: Contrasted contribution of intraseasonal time scales to surface chlorophyll variations in a bloom and an oligotrophic regime. Journal of Geophysical Research: Oceans, 125(5), e2019JC015701, 2020.

- Comparison of satellite and ESM timeseries: The analyses use satellite timeseries spanning 16 years and ESM simulations spanning 33 years. It would be helpful if the authors addressed whether this difference could impact the representation of multi-annual variability in the analyses and thereby affect their findings and conclusions.

This was previously insufficiently clear. We provided the comparison between satellite SChl and historical CMIP6 simulations for the common period of 1998–2014. To clarify this, we will revise line 110 by adding the sentence: 'The comparison between satellite observations and CMIP6 historical simulations is provided for the common period, 1998–2014.'

The CMIP6 data for 1981–1997 is used exclusively for Figure 9, to explore whether the changes observed between the two periods (1998-2014 and 2084-2100) can be attributed to decadal variability.

- Thresholds for spatial coherence analysis: When mentioning the thresholds for the analyses of the spatial extent of coherence, the authors could clarify the rationale behind the choice for an upper threshold of 2400 km in diameter and the threshold value of 0.8 for correlations and how these influence the findings and interpretations.

Our focus was on the spatial scales of subseasonal and multiannual variability, which typically occur at smaller scales, predominantly below 2000 km. Setting an upper threshold of 2400 km allowed us to concentrate on the relevant scales while reducing computational time and energy consumption.

The choice of a 0.8 correlation threshold was somewhat arbitrary, but it represents a high degree of spatial coherence, providing confidence in the robustness of identified patterns. Sensitivity tests conducted for the Mediterranean Sea (Supplementary Figure 4 of Keerthi et al., 2020), varying the threshold between 0.75 and 0.85, showed that our results were only weakly sensitive to changes within this range. This indicates that the findings and interpretations remain consistent across slightly different correlation thresholds.

- ESM future simulations: It could be mentioned why the future simulations analysis was limited from 2084 to 2100, rather than a longer time range.

We used a consistent 16-year time span for all analyses, including historical and future simulations. The satellite SChl observations and historical simulations only share a 16-year common period (1998–2014).

- Use of a single ensemble member: The study currently uses one single ensemble member per ESM. It would be interesting to discuss the implications of this choice, as utilizing the ensemble mean could provide a more accurate representation of model

performance and reduce variability introduced by individual simulations. Similarly, where possible, it would be valuable to discuss the mean across ESMs, as ensemble means often yield more accurate representations than individual models.

We agree with the reviewer that ensemble means often provide more accurate representations than individual models. However, the primary objective of this study is to evaluate the ability of each model to simulate temporal variability in SChl, rather than to identify the best-performing model. This approach is intended to offer insights that modeling groups can use to enhance their models further.

**Technical corrections:**

Ln 81: Umlaut on Müller

Corrected

Ln 84: …more than 'three' times...

Corrected the sentence ' - MPI-ESM1.2-HR has a horizontal resolution twice as high for the atmospheric component (100 km) and more than three times as high for the oceanic component (~40 km) compared to MPI-ESM1.2-LR (200 km and 150 km for the atmospheric and oceanic components, respectively)'.

Ln 88: Keerthi et al. (2022) (comma is not necessary)

Corrected – 'We utilised the datasets outlined in Keerthi et al. (2022) for observed SChl and SST. The SChl data is the Level 3 Mapped 9x9 km resolution 8-day averaged product (release 4.1), covering the period from January 1998 to December 2014'.

Table 1: A  border line is missing between HAMOCC6 and 150 km

Corrected

| CMIP6 Simulations | Physical Ocean Model | Ocean BGC Model | Horizontal resolution (Physical & BGC Model) | Model Simulations | References |
|---|---|---|---|---|---|
| IPSL-CM6A-LR | NEMO-OPA | PISCES | 100 km | Historical | Boucher et al. 2018, 2021; Séférian, 2018 |
| IPSL-CM6A-LR-INCA | | | | | |
| CNRM-ESM2-1 | | | | | |
| CESM2 | POP2 | MARBL | 100 km | Historical | Danabasoglu, 2019a, b, c; |
| CESM2-FV2 | | | | | |
| CESM2-WACCM-FV2 | | | | | |
| MPI-ESM1.2-HAM | MPIOM | HAMOCC6 | 150 km | Historical | Neubauer et al., 2019 ; Wieners et al., 2019a, b, c; Jungclaus et al., 2019a, b; Schupfner et al., 2019 |
| MPI-ESM1.2-LR | | | | Historical, SSP5-8.5, piControl | |
| MPI-ESM1.2-HR | | | 40 km | | |
| NorESM2-LM | MICOM | HAMOCC | 100 km | Historical, SSP5-8.5 piControl | Seland et al., 2019a, b, c ; Bentsen et al., 2019a, b |
| NorESM2-MM | | | | Historical, SSP5-8.5 | |

**Table 1.** The CMIP6 Earth system models used in this study; their individual components used to represent ocean and marine biogeochemistry; nominal horizontal resolutions of their ocean and marine biogeochemical models; simulations that were assessed.

Ln 126: Reference the CDO remapping tool remapdis (see reference on: https://code.mpimet.mpg.de/projects/cdo/wiki/Cite)

Thank you for noting it. The citation will be added .Schulzweida, Uwe. (2023). CDO User Guide (2.3.0). Zenodo.https://doi.org/10.5281/zenodo.10020800

Ln 173: 'display' in plural instead of displays

The sentence has been modified – 'CESM2, CNRMESM2-1, and IPSL-CM6A-LR display varying biases relative to satellite SChl across regions'.

Ln 182: Perhaps mention the metric of correlation employed, I assume the Pearson Correlation coefficient?

The sentence has been modified –'The spatial correlation (pearson correlation) between CMIP6 models and observations remains below 0.6 (Fig. 2a), with MPI models showing particularly low correlations, below 0.2'.

Fig 2: Add degree symbol at 60°N and 60°S. For clarity, consider adding to the description that the dots represent models, while dashed lines represent observations. Additionally, complement the color scheme by using different symbols for each model to improve accessibility for color-blind readers.

The Figure 2 and caption has been modified.

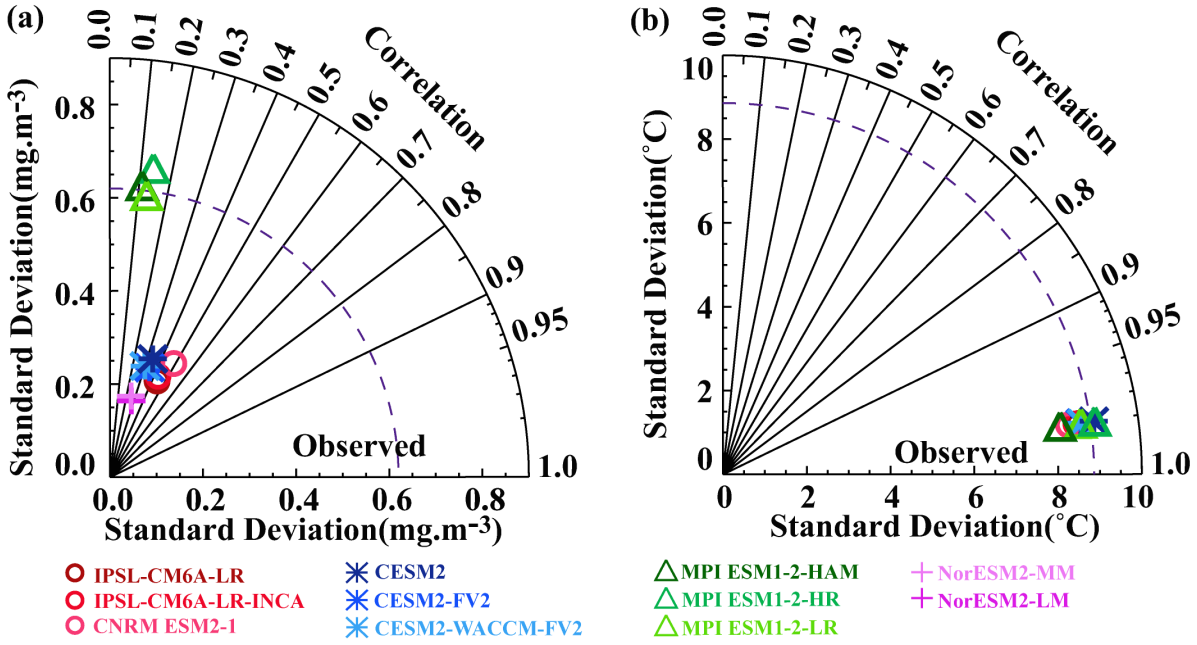


Figure 2: Evaluation of the mean spatial distribution. Taylor diagram for the annual mean (a) SChl and (b) SST over the period 1998–2014, within the domain 60°N–60°S. The dashed curve represents the standard deviation of the observational data.

Ln 216: There is no reference to Figure 3 in the text. It would improve clarity to reference Figure 3 here.

The sentence will be modified to 'The variability of SChl across different timescales varies significantly among the CMIP6 simulations (Figure 3)'.

Fig 3: The description states '(Left Panel)', however, I do not see a left and right panel nor a reference to a '(Right Panel)'. Consider adding for clarity: Normalized standard deviation 'of globally averaged' Schla…

Sorry for the confusion. This will be corrected. In the initial stage of the manuscript, Figures 3 and 4 are combined as the left and right panels.

Ln 266-267: Consider adding a reference to the figures in the sentence: The standard deviation across different timescales and the relative contribution of these timescales to the total SST variance '(Figure 4)' show distinct patterns compared to SChl '(Figure 3)'.

It will be corrected as suggested by the reviewer. Thank you for noting it.

Ln 273: The term 'ENSO' is used as an abbreviation without prior introduction. Additionally, 'El Niño' is mentioned in line 469. For coherence and clarity, consider introducing the term in full as 'El Niño–Southern Oscillation' upon its first use, then consistently using either 'El Niño' or 'ENSO' throughout the rest of the manuscript.

We will modify line 273 to introduce ENSO as "El Niño–Southern Oscillation" and will change the term "El Niño" in line 469 to ENSO.

Description Fig 4: Is it standard practice to reference a previous figure or would it be helpful to include the full description again?

Figure 4 caption will be modified as shown below.

Figure 4: Variability across timescales for SST: Similar to Figure 3, but for SST. (Left Panel) (a) Normalised standard deviation of SST from observations and CMIP6 historical simulations. Standard deviation at each grid point is normalised by the mean over each grid. (b) Percentage of SST variance explained by each component (sub-seasonal, seasonal and multiannual) for observations and CMIP6 historical simulations. Shading represents the different model groups described in Section 3.2, with green for Group 1, pink for Group 2, and blue for Group 3. Note that NorESM2-LM and NorESM2-MM are excluded, as daily resolution SST data for these models is not available on the CMIP6 data portal.

Fig 6 ln 353: 800 km in lowercase

Corrected

Ln 473-474: Consider adding the following: The simulated change of the sub-seasonal variability of SChl 'in response to X',…

The following sentence will be modified as 'Specifically, the simulated impact of climate change on the sub-seasonal variability of SChl has, to the best of our knowledge, not been previously assessed in CMIP-type models.