We thank the reviewers for the time taken to review this work and for their insightful comments. Please see the responses below in blue.

Reviewer report #1

This study extracts indices about phytoplankton bloom from the ocean color dataset. As the authors pointed out, this approach has been used by several communities. Previous studies are well revisited; three different methods are adopted and cross-compared. Details about the method shown in the manuscript, that may help the other researchers who want to develop more advanced techniques. The indices are helpful to understand ecosystems and frequently adopted by many studies. Nonetheless, it was seldom provided as a dataset. In my opinion, this study and the datasets may be worth publishing in the Earth System Science Data once several concerns (especially those for Figure 4) are resolved. Below are my comments:

Figure 4 is wired for me. Figure 4b shows bloom duration longer than a year (larger than 400 days), that is not realistic and does not make sense for me. The SCR, that is a correlation (based on the definition stated in the manuscript), cannot exceed 1 (100%), nevertheless Figure 4c shows 1<SCR. The peak at SCR=1 (Figure 4c) is nonsense for me too. I presumed that some process in Figure 4 is not mentioned, or something is wrong here.

We thank the reviewer for highlighting these important concerns regarding Figure 4.

With regards to bloom duration longer than a year. To note, our bloom detection method does not constrain or force a bloom slice to be within a 12 month period, as is done in some other phenology studies e.g. Henson et al. 2018: *""For each calendar year we first identified the date of peak chlorophyll concentration and then concatenated the preceding and following 6 months."* So in their method a bloom can never be longer than 12 months. Instead, in our method, we recognise that there are areas where blooms can have multiple peaks. *"The 'bloom slice': The bloom slice, used to find the bloom initiation and termination dates, is identified for each pixel as the 6-month time span preceding and following from the maximum bloom peak (ii). Or in the case of multi-modal blooms, 6-months preceding the first and following the last peak respectively."* Typically blooms that have longer durations are found in oligotrophic gyres, characterised with low SCR (weak seasonality), highly variable and poorly defined blooms.

We have added the following text to clarify further, see section 3.1 in the manuscript:
In these oligotrophic regions, where bloom amplitude is constrained by nutrients, the seasonality of phytoplankton blooms is not well-defined and characterised by high intraseasonal variability (Figure 2, Thomalla et al., 2011). Worth noting when applying our bloom detection method to these regions is that it does not constrain a bloom slice to be within a 12 month period, as is done in other phenology studies (e.g. Henson et al., 2018). Rather, by allowing for multiple peaks to be considered within a bloom, this approach may produce extended bloom durations that are beyond a year in regions with no discernable or strongly defined seasonal cycle.

Furthermore, we have also added a new section: "***4 Limitations of the phenology algorithm and future developments"*** which highlights the limitations of such an approach, the detection methodology and future developments.

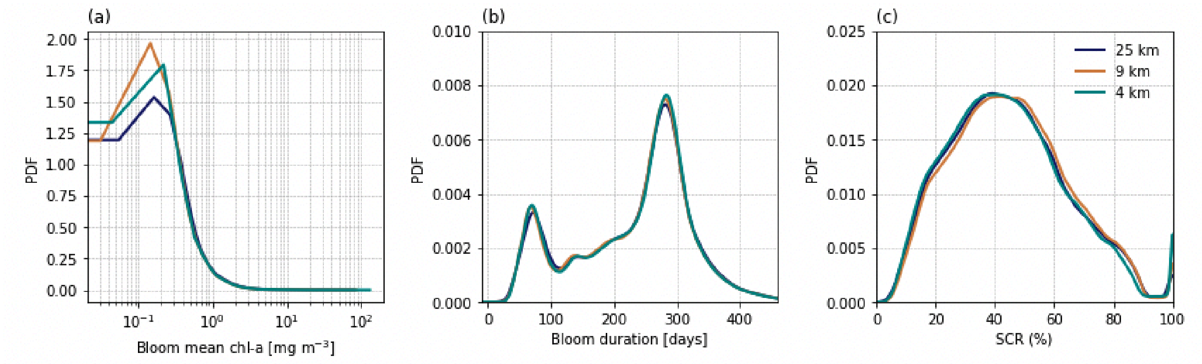
Regarding the SCR values that are 1. Over the entire dataset, all years, this only occurs in ~0.20% of the data. The reason for this is floating point errors when performing the regression between the climatological time series and the time series for each year. Below is a screenshot showing the maximum value of SCR, where the floating point errors manifest at the 16th decimal place.

```
print(f"{scr.max().values.item():.32f}")
```

```
1.00000000000000088817841970012523
```

To avoid any confusion we have removed any values where this floating point error occurs in the dataset.

Please see revised Figure 4. We have put the log-scale on the x-axis and unconstrained the y-axis for Figure 4 (a).



I would like to suggest using “satellite-driven” rather than “observational” in the entire manuscript including the title, because it can be easily misunderstood the data set using the in-situ observations. As the authors may know, satellite-driven measurements are occasionally not considered as observations due to the issues mentioned by the authors (gaps in the measurements and errors including bias in the algorithm). I think that “satellite-driven” is more clearly state the products in this study.

We have changed “observational” to “satellite-derived” throughout the text including the title of the manuscript. We chose satellite-derived over satellite-driven, we hope the reviewer agrees with this choice.

Minor comments:
L64: Typo? Not “Quay, 2017) Having” but “Quay, 2017). Having”.
Noted and corrected. Thank you.

L158: Feel like that the abbreviation “SO” never been stated before. I presumed that it stands for Southern Ocean and suggest that do not use abbreviation.
Thank you for spotting this. We have removed the abbreviation and replaced with Southern Ocean

Figure 4a: Entire PDF (including the peak) for Bloom mean chl-a needs to be shown or, at least, stated. Log-scale axis or stating the information about the peak in caption may be helpful.
Updated figure 4. Please see above.

Figure A1: Is the time series from the in-situ observations provided by the stations or from the satellite measurement at the location of stations? This should be stated in either the caption of figure or the manuscript (maybe near L468).

We have updated the caption of Figure A1 to include: "Figure A1: Examples of phytoplankton bloom seasonal cycles of ***satellite-derived chlorophyll-a from*** *OC-CCI* and comparisons in phenological detection methods at….."

Reviewer report #2

This manuscript presents a satellite-derived chlorophyll-a dataset from the Ocean Colour Climate Change Initiative, providing phenological metrics at 4, 9, and 25 km spatial resolutions. The dataset is accessible and can be analysed easily using GIS/coding. The dataset is highly valuable for various research applications, including ecosystem monitoring, biodiversity assessments, and climate impact studies. The study is well conceived and has the potential to make a significant contribution to the field. Below, I provide some minor comments and suggestions for improvement: 1. A key concern is the lack of comparisons with prior phenology studies, particularly those utilizing in situ observations. While ship-based in situ measurements are indeed limited in spatial and temporal coverage, autonomous platforms such as BGC-Argo floats offer continuous data that could be utilized for such comparisons. Including a case study or analysis demonstrating agreement between satellite-derived and in situ-derived phenology metrics would enhance the dataset's credibility and highlight its utility. Additionally, discussing how the proposed dataset aligns with or diverges from earlier findings could provide valuable context. For reference, here are some relevant studies that might inform such comparisons (no need to cite; they are provided for your consideration):

For reference, here are some relevant studies that might inform such comparisons (no need to cite; they are provided for your consideration):

Demetriou, M., Raitsos, D. E., Kournopoulou, A., Mandalakis, M., Sfenthourakis, S., & Psarra, S. (2022). Phytoplankton Phenology in the Coastal Zone of Cyprus, Based on Remote Sensing and In Situ Observations. Remote Sensing, 14(1), 1–16. https://doi.org/https://doi.org/10.3390/rs14010012

Gittings, J. A., Raitsos, D. E., Kheireddine, M., Racault, M. F., Claustre, H., & Hoteit, I. (2019). Evaluating tropical phytoplankton phenology metrics using contemporary tools. Scientific Reports, 9(1), 1–9. https://doi.org/10.1038/s41598-018-37370-4

Kalloniati, K., Christou, E. D., Kournopoulou, A., Gittings, J. A., Theodorou, I., Zervoudaki, S., & Raitsos, D. E. (2023). Long-term warming and human-induced plankton shifts at a coastal Eastern Mediterranean site. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-48254-7

Kournopoulou, A., Kikaki, K., Varkitzi, I., Psarra, S., Assimakopoulou, G., Karantzalos, K., & Raitsos, D. E. (2024). Atlas of phytoplankton phenology indices in selected Eastern Mediterranean marine ecosystems. Scientific Reports, 14(1), 9975. https://doi.org/10.1038/s41598-024-60792-2

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

We thank the reviewer for this great suggestion, and have added an additional paragraph which uses some of the examples provided above and others to make some regional comparisons with the data produced presented here and other phenology studies. See additional text (section 3.1):

"A comparison of our satellite-derived phenology product with bloom indices derived from in situ data at a selection of regional case studies shows reasonable agreement. For example, in the Saronikos Gulf (Eastern Mediterranean), Kalloniati et al. (2023) report a mean bloom initiation in early October (2005–2015), which

compares well with our mean bloom initiation over the same period of 24 September. Similarly, their mean bloom peak occurs in late February, closely matching our estimate of 24 of February. However, there are notable differences in bloom termination with their approach reporting a seasonal bloom that terminates in mid-April, compared to our estimate of ~100 days later on 13 July. This discrepancy likely arises because their method does not account for multiple bloom peaks, whereas our method is specifically designed to include the secondary peak observed in April as part of the seasonal bloom (see their Figure 3c). Another example from long-term mooring observations (1998-2022) in the Bering Sea shelf (Nielsen et al. 2023) reports the timing of the bloom maximum to range annually between the end of April to mid-June (see their Figure 2), which compares well with our mean estimate over the same period of 25 of May (standard deviation of 57 days). In a Red Sea comparison, although our satellite derived phenology data product was able to detect similar bloom initiation and max peak timing for the primary bloom in winter (as observed by Racault et al., 2015), it is not designed to provide indices fort bi-modal blooms and thus is unable to identify the secondary bloom in summer. Beyond these existing studies, we applied our phenological detection method (TS) to chlorophyll-a data from the HOT and BATS long-term monitoring sites (Figure 2A, Valente et al. 2022). At HOT (1998-2018)(Figure 2Aa), the in situ bloom initiation occurred on 25 July (±48 days) compared to the satellite-derived occurring on the 21 July (±42 days), in situ bloom max timing on 12th of December vs. 5th of December, and termination on 22 May (±32 days) vs. 6 June (±29 days) and duration in situ of 299 days vs durations of 303 days from satellite data. Similar agreement was seen in the BATS station (Figure 2Ab).

In addition, we have used chlorophyll HPLC data from https://doi.pangaea.de/10.1594/PANGAEA.941318 to do some of our own in situ comparisons at the long-term monitoring sites HOT and BATS. See additional Figure A2 below:

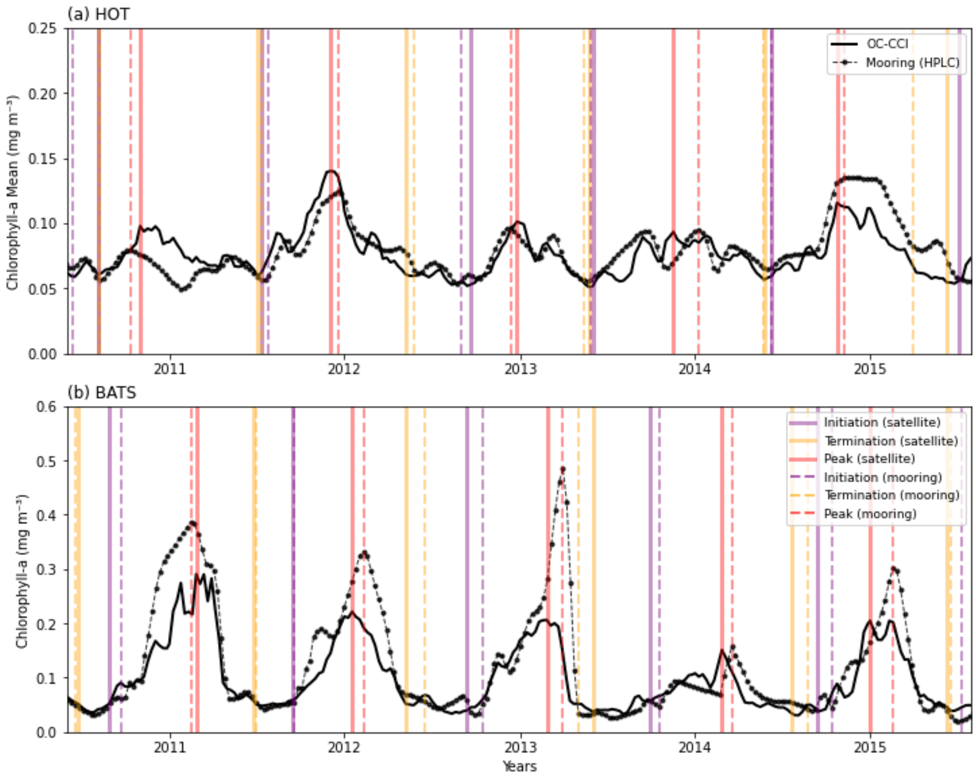
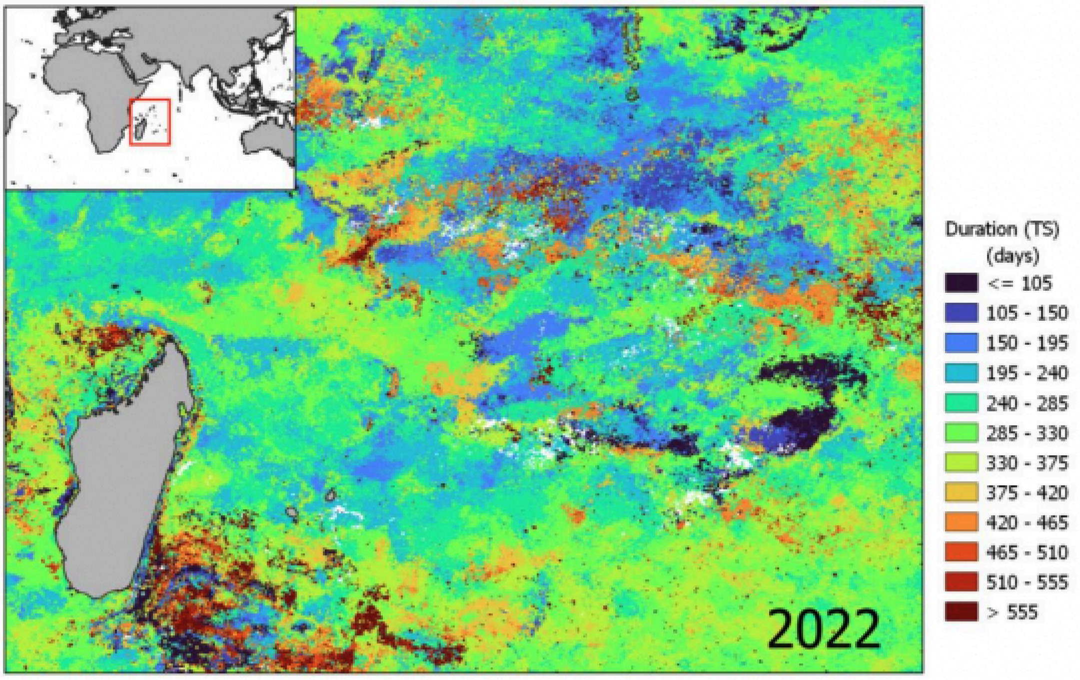


Figure A2: Comparison of five years of in situ chlorophyll-a measurements (Valente et al. 2022) with satellite-derived chlorophyll-a (OC-CCI), along with key phenological indices (solid and dashed vertical lines for satellite and mooring, respectively) at two sustained observing stations: (a) Hawaii Ocean Time-Series (HOT, 21° 20.6'N, 158° 16.4'W) and (b) Bermuda Atlantic Time-Series Study (BATS, 31° 50'N, 64° 10'W).

2. Lines 498–501: Creating temporal composites is important for mitigating potential noise caused by interpolation errors in the OC-CCI dataset. While the product is highly valuable, it does exhibit some irregularities that can influence the calculation of phenology metrics, particularly at higher resolutions (e.g., 4 km), and therefore impact the derived phenology indicators described in this study. For example, when examining the spatial and temporal variability of phytoplankton growth period durations in the Indian Ocean (Figure 1), several regions (pixels) display durations ranging from approximately 100 to 600 days over the years. While such variability could be realistic in certain cases, it is important to recognize that, like any dataset, this product includes some irregularities that may affect its outputs.



We thank the reviewer for highlighting this important point. We have included the following additional section, which notes irregularities and limitations of the data product. As well as future developments for new releases of this product.

4 Limitations of the phenology algorithm and future developments

The diversity of the phytoplankton seasonal cycles across the global ocean makes it challenging to generalise a single methodological approach that is capable of capturing all phenological metrics accurately. Our attempt to do so with this data product may lead to some irregularities, most notably when applied to regions with a poorly defined or unique seasonal cycle. For example, in ultra-oligotrophic regions where the bloom amplitude is particularly low and intraseasonal variability particularly high, our detection method prescribes long bloom durations that may exceed one year and can lead to overlapping bloom slices. Another example is regions with bi-modal blooms, where there is a well-defined summer and winter bloom in a given annual cycle. Although our phytoplankton phenology detection method is designed to allow for multiple peaks to occur within a bloom cycle; it has not been designed to cater for bimodal annual cycles, which would require the identification of separate summer and winter initiation and termination indices. In these instances our method may result in extended bloom durations. While these regions are relatively uncommon (e.g. Racault et al 2017, Figure 2c), they do exist, as is the case with the Red Sea (Racault et al. 2015). Future developments of this data product will endeavour to incorporate updates and improvements to the detection methods to better cope with these irregularities. We welcome users to reach out if other irregularities are identified within a specific area of *interest and to work with the authors to improve future versions of the product. All future changes to the product will be fully documented on Zenodo as new versions are released.*