Assessment of gap-filling techniques applied to satellite phytoplankton composition products for the Atlantic Ocean

Ehsan Mehdipour[1,2], Hongyan Xi[1], Alexander Barth[3], Aida Alvera-Azcárate[3], Adalbert Wilhelm[2], Astrid Bracher[1,4]

[1]Alfred Wegener Institute (AWI), Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]School of Business, Social & Decision Sciences, Constructor University, Bremen, Germany
[3]GeoHydrodynamics and Environment Research (GHER), University of Liège, Liège, Belgium
[4]Institute of Environmental Physics, University of Bremen, Bremen, Germany

Correspondence to: Ehsan Mehdipour (ehsan.mehdipour@awi.de)

Author Comments in response to Referee #1

This manuscript presents a timely and relevant study that assesses the performance of two gap-filling methods, DINEOF and DINCAE, applied to satellite-derived chlorophyll-a and phytoplankton functional type products in the Atlantic Ocean. The study is well-motivated, and the methodological rigor is commendable. The work contributes meaningfully to the ocean color remote sensing community and addresses a practical challenge in producing spatially continuous biogeochemical products. Therefore, I recommend this paper for publication. However, there are several aspects of the manuscript that would benefit from revision to improve clarity, structure, and academic presentation.

- We sincerely thank the reviewer for the constructive and detailed comments and suggestions, which have greatly helped us to improve the quality and clarity of the manuscript. We have carefully addressed each point in our responses below and will revise the manuscript accordingly, as indicated in the respective replies.

1. The manuscript is generally well-written, but there are a few areas where sentence structure and clarity could be improved. The introduction section is overly long and contains redundant background descriptions, which compromise the logical flow. It is recommended that the authors streamline this section by first clearly identifying the problem posed by data gaps in phytoplankton functional type products, followed by a focused overview of existing gap-filling methods, and then a concise rationale for selecting and comparing DINEOF and DINCAE in this study. A clearer structure will help readers better grasp the novelty and objectives of the work.

- We have restructured the Introduction section to improve clarity and logical flow. The revised version follows a three-step structure as suggested: (1) a clear identification of the problem related to data gaps in PFT products, (2) a concise overview of existing gap-filling approaches, and (3) a focused rationale for selecting and comparing DINEOF and DINCAE in this study. Redundant background information has been removed. The changes can be found on pages 2–4 of the revised manuscript.

2. The detailed explanation of the DINEOF and DINCAE methods is commendable, providing insight into their respective advantages and limitations. While the methods section is detailed and provides useful information for reproducibility, the level

of granularity is at times excessive, resembling a technical report rather than a scientific manuscript. Several parts describe operational or troubleshooting procedures in a narrative style that could be significantly condensed. For instance, the explanation regarding the use of the holdout method for generating the test dataset includes unnecessary implementation details that do not directly contribute to the scientific understanding. It is suggested to summarize such content more concisely, focusing on methodological rationale rather than process narration. Similarly, generic background statements—such as those explaining what hyperparameters are and why they matter—can be omitted or substantially shortened, with emphasis placed instead on the specific choices made in this study and their justification.

- We have carefully revised the Methods section to remove or condense narrative descriptions and operational details that were not essential for scientific understanding. Specifically, we shortened the explanation of the holdout method and eliminated generic background information (e.g., on hyperparameters) to focus more on the methodological rationale and the specific choices made in this study. As a result, the section is now more concise and maintains a clearer focus on reproducibility and relevance. The revisions can be found in Sections 2.2.3 and 2.3.3 of the revised manuscript.

3. The results are well-presented, particularly the comparisons between DINEOF and DINCAE gap-filling methods. However, some of the claims about the relative advantages of DINCAE over DINEOF, especially in complex regions, could be better substantiated with more detailed statistical analysis. For instance, the statement that DINCAE outperforms DINEOF in specific regions (e.g., coastal areas) could benefit from a deeper analysis of the underlying causes for these discrepancies. While the manuscript effectively outlines the strengths and limitations of both gap-filling methods, it could further benefit from a more detailed discussion on the future improvements or modifications required for both techniques. Potential solutions to optimize computational efficiency could be discussed.

- The claim that DINCAE had better results in the complex region was thoroughly justified in section 3.2 (Performance evaluation) by showing lower errors in Figures 5 and 6, particularly in these regions. Section 3.3.1 (Gradient field) also showed the better reconstruction of gradients in the dataset. This claim does not mean that the DINCAE model has a thoroughly better performance than DINEOF, as validation with in situ measurements shows better results for DINEOF. We have toned down the statement of comparison between DINCAE and DINEOF to make sure that the transferred message is compatible with the statistics (see lines 24, 408, 578, 588, 645, 755, 765 of the revised manuscript).
- The following text was added in Section 3.2 to explain the underlying causes for the discrepancies between DINEOF and DINCAE in dynamic regions (see lines 471–477 of the revised manuscript).
    - "The discrepancy between the two models in dynamic regions may arise from their fundamental methodological differences. DINEOF reconstructs missing data by extracting dominant spatiotemporal modes from the entire temporal domain, with an additional emphasis on the local spatiotemporal structure through a Laplacian filter. This EOF-based reconstruction can underrepresent transient or localized features. In contrast, DINCAE employs a U-Net-style architecture that interpolates missing values based on nearby spatiotemporal information, allowing it to more effectively capture localized or transient variability in the data by preserving fine-scale details through skip connections."
- The future improvements or modifications required for both techniques are explained extensively in Section 3.5 (Novelty and Limitations). The following recommendation was added to state the potential solutions to optimize computational efficiency (see lines 694–700 of the revised manuscript).

- “The computational efficiency of both models can be enhanced by introducing an internal data segmentation step prior to pattern extraction, which would allow parallelized computations across multiple clusters or nodes. For DINEOF, this approach could reduce the cost of iterative EOF decomposition by processing spatial segments independently and later merging the reconstructed fields. For DINCAE, parallelization can be achieved by distributing training and inference over spatial subsets or by adopting model architectures optimized for distributed GPU computation. Additionally, implementing chunked data handling and memory-efficient input/output could further optimize large-scale processing for both methods.”

4. The study provides an important contribution to gap-filling in marine remote sensing, but the broader environmental implications of the findings could be more thoroughly explored. Specifically, how might these improved gap-filling techniques contribute to more accurate global phytoplankton monitoring, climate change studies, or marine conservation efforts?

- To better highlight the broader environmental implications of our findings, we included the following statements in Section 4 (Conclusion and outlook) of the revised manuscript (see lines 732–740 and 778–781):
  - “Missing data in satellite-derived biogeochemical observations can lead to underrepresentation of important spatiotemporal dynamics, especially in regions characterized by high variability or ecological sensitivity, such as coastal zones and upwelling areas. Missing data in these regions can obscure critical information about transient biological events and environmental responses. The application of robust gap-filling techniques enables the reconstruction of these dynamic patterns, providing more complete datasets that can be exploited for improved modelling, targeted field campaigns, and continuous environmental monitoring. For example, enhanced reconstructions of Chla concentration can support fisheries management by linking biological productivity with fish distribution, assist in the early detection and prediction of harmful algal blooms, and improve estimates of net primary production. These downstream applications ultimately contribute to the development of more sustainable marine and climate management strategies.”
  - “Although the current methods are best suited for regional monitoring, scaling them to global applications would require additional optimization. Nevertheless, constructing a globally gap-filled Chla dataset, even at a reduced spatiotemporal resolution, could provide invaluable input for long-term climate assessments, global biogeochemical modelling, and validation of Earth system models.”

Assessment of gap-filling techniques applied to satellite phytoplankton composition products for the Atlantic Ocean

Ehsan Mehdipour[1,2], Hongyan Xi[1], Alexander Barth[3], Aida Alvera-Azcárate[3], Adalbert Wilhelm[2], Astrid Bracher[1,4]

[1]Alfred Wegener Institute (AWI), Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]School of Business, Social & Decision Sciences, Constructor University, Bremen, Germany
[3]GeoHydrodynamics and Environment Research (GHER), University of Liège, Liège, Belgium
[4]Institute of Environmental Physics, University of Bremen, Bremen, Germany

Correspondence to: Ehsan Mehdipour (ehsan.mehdipour@awi.de)

Author Comments in response to Referee #2

The study entitled "Assessment of gap-filling techniques applied to satellite phytoplankton composition products for the Atlantic Ocean" by Mehdipour et al. addresses the critical challenge of missing data in satellite-derived ocean color products, particularly phytoplankton functional types (PFTs), which are crucial for understanding marine biogeochemical cycles and climate change impacts. The authors evaluate two established gap-filling methods, DINEOF and DINCAE, for reconstructing these datasets in the Atlantic Ocean. The manuscript is well written, interesting to read, the methodology sounds valid, and results and conclusions seem to be reasonable. I only have minor comments and suggestions for further improving the manuscript.

- We sincerely thank the reviewer for the thoughtful and constructive feedback, which provided valuable insights and helped us further improve the manuscript. We have carefully considered all comments and suggestions and have addressed each point in detail below. The corresponding revisions will be incorporated into the manuscript as indicated in our responses.

1. Page 8: the last two paragraphs are exactly the same (repeated).

- The duplicated paragraphs on page 8 have been removed.

2. DINEOF and DINCAE are different in terms of their input data. For DINCAE, only three days of data (previous and next day) are used to reconstruct the missing pixels. However, the DINEOF processing needs data of much longer time series as input. From Section 2.3.1, it is not clear how many days of data are used for input. I suggest that this important information included in the description. Also, please discuss the effect of the length of the time series on the performance of the two models.

- The number of days of data used for reconstruction is indeed a hyperparameter that we tuned during model development. In DINEOF, although the full dataset is available for extracting EOF values, the effective temporal covariance is influenced by the Laplacian filtering parameters (α and P). The extent of this filter, calculated as $L = 2\pi\sqrt{\alpha p}$, was approximately 5.74 days in our final model, indicating that temporally close datasets (within about six days) have the greatest influence during reconstruction (see Section 2.3.3). For DINCAE, several window sizes (3, 5, and 7 days) were tested during hyperparameter tuning of the ntime_win parameter. The model using a 3-day window achieved the best cross-validation performance and was therefore selected for the final reconstruction. These details are already included in the text.

- A new statement has been included in Section 3.5 (Novelty and limitations) discussing the impact of time series length on the performance of the two models (see lines 692-701 of the revised manuscript).
  - "Furthermore, the length of the time series can significantly influence the performance of both models. A short time series may result in EOF patterns that do not fully capture the underlying variability in the dataset when using the DINEOF method, and it may also lead to suboptimal tuning of parameters and hyperparameters in DINCAE due to limited training samples. A long time series primarily affects the DINEOF method, as EOF extraction over extended periods tends to emphasize more persistent and large-scale spatial patterns while reducing the representation of transient variability. This effect is less pronounced in DINCAE, since the data are temporally segmented into minibatches before being processed by the network, which decreases, but does not eliminate, its dependence on the total record length. When the time series is extended, the corresponding increase in training epochs allows the model to learn from a broader range of examples and improves its generalization ability, but it may also lead to a reduced sensitivity to short-lived or rare features. Although not examined in the present study, the length of the time series could be considered an experimental hyperparameter to be optimized during model development."

3. Equation (1) is confusing: t is not defined, SST is not included, T is not specified, X-CHL is repeated for T times?

- We have reformulated Equation (1) to improve clarity. All parameters, including t and T, are now explicitly defined in the text, and SST has been incorporated into the formulation. Additionally, the notation X_{CHL}^{1} now clearly represents the total chlorophyll-a concentration dataset for day 1. The revised equation and its description can be found in Section 2.3.1 of the manuscript.

4. Section 3.2 and Fig. 5-6: Please discuss the effect of the data miss rate on the model performance in terms of different parameters (TChl vs. PFTs), and spatial regions (high latitude, equatorial Vs. mid-latitude).

- It is challenging to directly compare the effect of data availability across different spatial regions (e.g., high-latitude, equatorial, and mid-latitude areas), since the physical and biological dynamics in these regions are inherently distinct. This variability makes it difficult to isolate the impact of data availability from the influence of regional dynamics and create a one-to-one comparison. To investigate the relationship between data availability and model performance across different groups, we generated 3D histograms showing the relationship between data availability and the mean absolute error (MAE) for each pixel in the test dataset (Figures S3 and S4, shown at the end of this document). These plots illustrate how model performance (in terms of MAE) varies with changes in data availability. Both models show that higher data availability generally corresponds to lower and more tightly clustered MAE values, indicating improved reconstruction accuracy. In particular, the MAE for TChla decreases as data availability increases. When comparing the two models, DINCAE consistently produces reconstructions with lower MAE than DINEOF, suggesting that DINCAE more effectively captures the underlying spatio-temporal variability, especially in regions with limited data availability.
- These figures were not added to the manuscript, but a short statement referring to this work was added:

- “In addition, it is challenging to isolate the effect of data availability on model performance across different spatial regions (e.g., high-latitude, equatorial, and mid-latitude areas), since the physical and biological dynamics in these regions are inherently distinct. The investigation of the relationship between data availability and gap-filling model performance across different groups based on mean absolute error (MAE) (results not shown) showed, for both models, as expected, that a higher data availability generally corresponds to lower and more tightly clustered MAE values, indicating improved reconstruction accuracy. When comparing the two models, DINCAE consistently produces reconstructions with lower MAE than DINEOF, suggesting that DINCAE more effectively captures the underlying spatiotemporal variability, especially in regions with limited data availability. “

5. Page 23, line 557: change VIIRS-SNAP to VIIRS-SNPP

- The term “VIIRS-SNAP” has been corrected to “VIIRS-SNPP” in the revised manuscript.

6. It is noted that the PFTs shows significant difference with in situ measurement, but they are still valuable datasets. Further improvement of the quality of the PFT data are expected in the future.

- We agree that although the current PFT and gap-filled PFT products show discrepancies compared to in situ measurements, they remain valuable for large-scale and long-term analyses. We have emphasized this point in the conclusion section, highlighting the limitations of the current PFT datasets and the need for future improvements in retrieval accuracy and data quality. The revised text can be found in Section 4, Conclusion and Outlook, and the following statement was added:
    - “Although the current methods are best suited for regional monitoring, scaling them to global applications would require additional optimization. Nevertheless, constructing a globally gap-filled Chla dataset, even at a reduced spatiotemporal resolution, could provide invaluable input for long-term climate assessments, global biogeochemical modeling, and validation of Earth system models.”

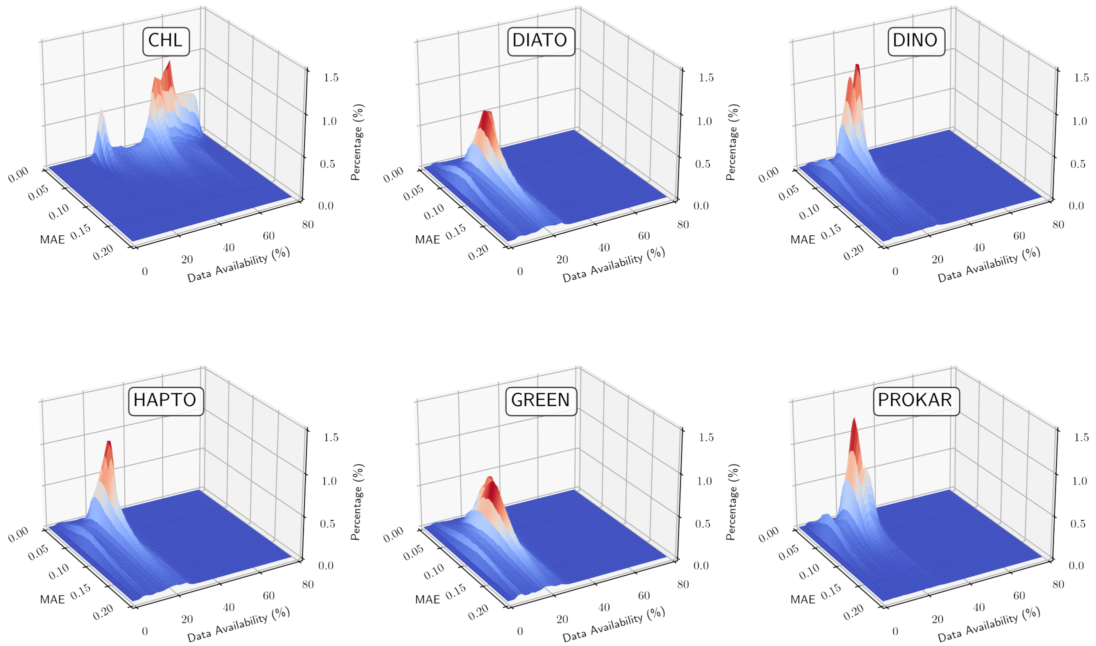

Figure S3. Histogram distribution of the test dataset's MAE vs Data Availability for DINCAE gap-filling

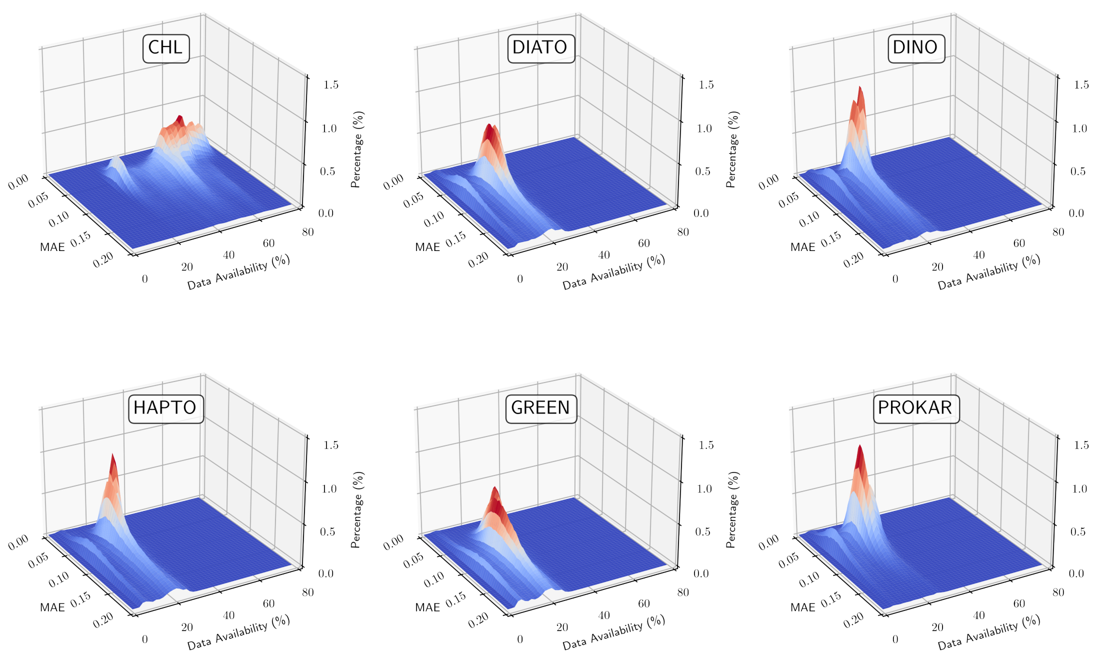

Figure S4. Histogram distribution of the test dataset's MAE vs Data Availability for DINEOF gap-filling