To Reviewer 1

First, we appreciate the reviewer's valuable comments. For your comments, we gave our corresponding explanations and responses below:

1. Page 1 Line 38: "satellite-derived missing data reconstruction is receiving increasing attention" The phrasing can be improved. Suggest: "reconstruction of missing satellite-derived data"

Response: This phrasing has already been modified to: "reconstruction of missing satellite-derived data is receiving increasing attention".

2. Page 4 Lines 21-23: Please also consider citing a paper with the detailed descriptions of current MODIS SST retrieval algorithm, e.g.: Jia, C., & Minnett, P. J. (2020). High latitude sea surface temperatures derived from MODIS infrared measurements. *Remote Sensing of Environment*, **251**, 112094.

Response: This reference has already been cited in the revised manuscript.

3. Page 5 Line 12: Firstly, since the resolution of MODIS data (4 km) is higher than that of AMSR2 (25 km), "downscaled" should be used instead of "upscaled" if the MODIS data were adjusted to the resolution of AMSR2 data. Otherwise, "upscaled" is correct. Secondly, have the authors considered the impact of this spatial aggregation as using nearest-neighbor

interpolation may introduce artifacts? Does that mean reconstructed datasets for all variables have the same spatial resolution? If so, can the authors make any comments on that considering the resolution of reconstructed MODIS data is lower in order to accommodate the AMSR2 data? A paragraph in the discussion section should be helpful.

Response: Here, “downscaled” is indeed appropriate, as the MODIS data were adjusted to match the AMSR2 resolution. This has been corrected to “downscaled” in the revised manuscript.

We acknowledge that the potential artifacts introduced by spatial aggregation were not explicitly considered. In some super-resolution reconstruction studies, nearest-neighbor or spline interpolation methods are commonly used to simulate low-resolution imagery. Following a similar approach, we downscaled the high-resolution MODIS data to match the lower resolution of AMSR2.

Previous studies have shown that resampling low-resolution data to a higher resolution often introduces larger errors, as this process is essentially an ill-posed (NS) problem. In contrast, downscaling high-resolution data to a lower resolution can significantly reduce the introduction of such errors. Based on this consideration, we chose to downscale the high-resolution MODIS data to the AMSR2 resolution, rather than upscale AMSR2 data to match MODIS.

This downscale process ensures that SST, SCHL, and SSW data are represented on a consistent spatiotemporal grid, allowing the construction of corresponding three-dimensional tensors. Without this resampling step, it would not be possible to unify datasets of different spatial resolutions into the same spatiotemporal framework.

A paragraph about this limitation has been added to the Discussion section: "Moreover, to unify MODIS SST and SCHL (high resolution) with AMSR2 SSW (low resolution), the nearest-neighbor method was used to downscale MODIS data to the AMSR2 resolution. Upscaling low-resolution data is an ill-posed (NS) problem that may introduce larger errors, whereas downscaling helps reduce error propagation. Therefore, MODIS SST and SCHL data were downscaled to the AMSR2 resolution rather than upscaling AMSR2 data to match the MODIS resolution. However, although the nearest-neighbor method is widely used for downscaling high-resolution imagery, it may still introduce artifacts that affect the reconstruction process. Consequently, adopting more advanced resampling techniques or using multi-source datasets with consistent spatial resolution could further improve reconstruction accuracy."

4. According to Table 1, the minimum SST in Subregion 3 is -1.81 ºC. Usually, a temperature threshold is applied to distinguish between ice (typically lower than -1.8°C) and open water. Even though this study

excludes high latitude regions, but a sea ice mask is still necessary for SST data but it seems not mentioned in the text.

Response: We re-examined the data. First, the values in Table 1 have been rounded; in reality, the minimum SST in subregion 3 is −1.8050 °C, differing from the ice threshold of −1.80 °C by only 0.005 °C. Second, statistics show that only one pixel falls below −1.80 °C. For the tensor in subregion 3, which contains 207,779 SST pixels, masking a single pixel would have a negligible effect on reconstruction accuracy. Considering this minimal difference from the ice threshold, we therefore did not introduce an ice mask in the revised manuscript. We hope the reviewer will understand and appreciate our rationale.

5. Page 6 Line 8: In Southern Hemisphere, the summer/winter months are opposite to the Northern Hemisphere. Please revise the descriptions, not only in this sentence, but also in Line 15 and Line 16 (Subregions 1 and 3 have the same seasonal pattern, peaking in the winter).

Response: In the revised manuscript, we have updated the relevant descriptions to use specific months directly, rather than seasonal terms such as summer or winter.

For example: Typically, the proportions of missing data are higher from June to August than in other months, especially in subregion 3.

6. Page 12 Line 20: Please also highlight the red ellipse in Fig. 7b as it states “both methods tend to underestimate high-value pixels”.

Response: In the revised manuscript, the corresponding red ellipse has been added in Fig. 7b.

7. Page 13: For Fig. 7, the color bar is not consistent with the density color shown in the plots. Same in Fig. 13.

Response: We have regenerated the color bars in Figs. 7 and 13 to ensure consistency with the density color shown in the plots.

8. For Figs. 9-11: Please do not consider the missing pixels as zero here because variables like SST could be 0 ºC causing unnecessary confusion (even though it is not the case for the northern Pacific in April), also because the difference between the reconstructed and original data is meaningless at those missing pixels. So, please set them as NaN and use another color for the missing pixels (e.g., gray) in the maps.

Response: In the revised manuscript, to distinguish them from the colors in the color bar, the missing pixels in the SST, SCHL, and SSW images are shown in black, while land pixels are shown in white.

9. Page 16 Line 8: Black ellipses are not found in Fig. 11d.

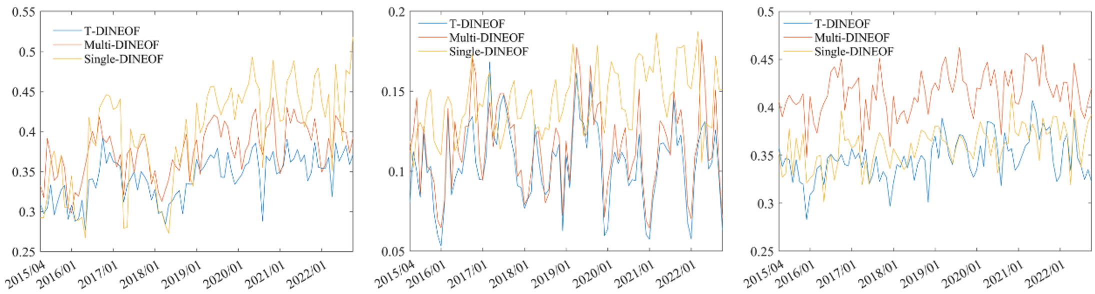
Response: A black ellipse has been added to Fig. 11d.

10. For the figure captions of Figs. 9-11, please add the information of the time of the map.

Response: The time information for the maps has been added to the figure captions of Figs. 9–11.

11. Page 20: For the temporal characteristics of the reconstruction accuracy, it might be useful to demonstrate the overall monthly RMSE variations during the experimental period to better reveal potential seasonal patterns.

Response: Due to space limitations, we have included the reconstruction accuracy results for SST, SCHL, and SSW over the entire study region in the Supplementary Materials (Fig. S3), and we hope for the reviewer's understanding. In the revised manuscript, we further analyzed the monthly RMSE variations over the entire study region and discussed the underlying causes of these patterns. The specific content is as follows: "Meanwhile, the reconstruction accuracy of SST, SCHL, and SSW over the entire study region was further evaluated (Fig. S3). The results show that, in most cases, T-DINEOF achieves the highest reconstruction accuracy, with particularly notable improvements for SST and SSW. For SCHL, however, the reconstruction accuracy of Multi-DINEOF is slightly lower than that of T-DINEOF. It is worth noting that for both SST and SCHL, the Single-DINEOF method exhibits the lowest reconstruction accuracy among the three methods, indicating that incorporating multivariate information can effectively improve reconstruction performance. In contrast, for SSW, the Multi-DINEOF method yields the lowest reconstruction accuracy. As discussed in Section 4.1, the correlation between SSW and the other variables (SST and SCHL) is relatively low. Therefore, the multivariate synergy in Multi-DINEOF does not enhance the reconstruction accuracy of SSW under low-correlation conditions. This also demonstrates that the T-DINEOF method, owing to its tensor-based reconstruction framework, is more effective in improving the reconstruction accuracy of variables with weak inter-variable correlations.”



(a) (b) (c)

Fig. S3 Monthly RMSEs for (a) SST, (b) SCHL, and (c) SSW over the entire study region.

12. Page 20 Lines 21-23: The monthly RMSE time series show interesting patterns, particularly the periodic behavior in subregion 1 for SCHL. The explanation linking this to homogeneity is plausible but could be strengthened with quantitative correlation analysis between RMSE and standard deviation.

Response: In the supplementary materials, we analyzed the standard deviation (std) of log(SCHL) across the three subregions to assess the spatial homogeneity of SCHL (Fig. S2). During winter, subregion 1 exhibits lower std values, indicating higher SCHL homogeneity. In general, more homogeneous regions tend to show lower reconstruction errors. Conversely, during summer, higher std values indicate greater spatial variability in SCHL, corresponding to increased reconstruction errors.

In subregion 3, the RMSE also shows a certain degree of periodic variation (Fig. 15f in the manuscript), but with smaller amplitude compared to subregion 1. This observation is consistent with the relatively lower std values of subregion 3 shown in Fig. S2.

Furthermore, based on the T-DINEOF method, we found that for log(SCHL) data, the correlation coefficient between RMSE and std reaches 0.65 in subregion 1, whereas it is only 0.02 and –0.14 in subregions 2 and 3, respectively, both significantly lower than in subregion 1. Therefore, we consider that the pronounced periodic variations in SCHL in subregion 1 significantly influence the variations in reconstruction accuracy, whereas in subregion 3, the smaller periodic variations in SCHL are only weakly correlated with changes in reconstruction accuracy.

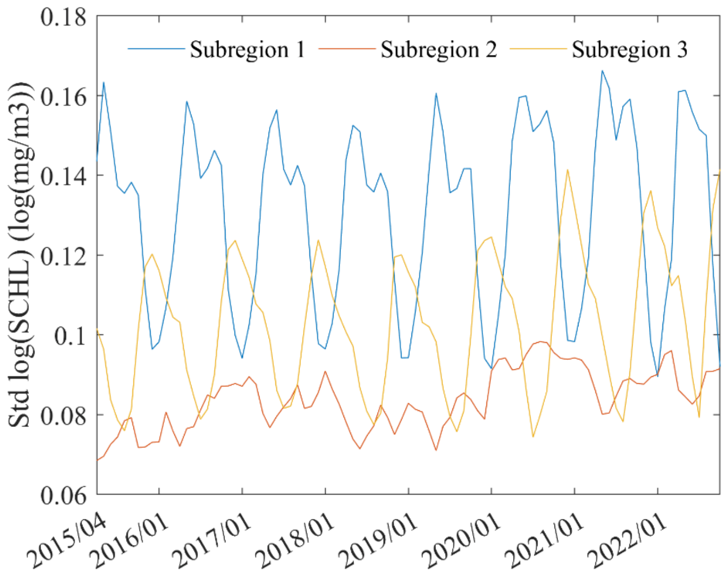


Fig. S2 Temporal std distribution (std) of log(SCHL) in the three study regions.

We have added the corresponding content in the revised manuscript as follows: "In addition, taking the T-DINEOF method as an example, the correlation coefficients between RMSE and std of log(SCHL) across the three subregions were calculated. In subregion 1, the correlation coefficient reaches 0.65, indicating that the pronounced periodic variations in SCHL significantly influence the variations in reconstruction accuracy. In contrast, in subregion 3, the smaller periodic variations in SCHL are only weakly correlated with changes in reconstruction accuracy (correlation = –0.14), and in subregion 2, due to the relatively small std values reflecting high SCHL homogeneity, the correlation between std and RMSE is negligible (correlation = 0.02)."

13. Page 22 Line 16: The authors acknowledge that T-DINEOF requires longer computation times, but no quantitative comparison is provided. Given that computational cost is a practical concern for operational applications, an added table comparing runtime for Single-DINEOF, Multi-DINEOF, and T-DINEOF across the three subregions is helpful.

Response: In this study, the T-DINEOF method was developed based on Matlab R2023b. During the algorithm development, we focused primarily on improving reconstruction accuracy rather than computational efficiency. Therefore, it should be acknowledged that there is room to optimize the execution speed of the code. Furthermore, tensor operations involved in T-DINEOF, such as T-SVD decomposition and tensor transposition, were independently implemented based on the tensor definitions. Steps such as data preprocessing, land masking, and selection of cross-validation pixels were also integrated into the execution code. As a result, it is currently difficult to provide absolute computation times for T-DINEOF.

In our experiments, we confirmed that T-DINEOF requires longer computation times compared with matrix-based methods, even though they have similar computational complexity, likely because tensor operations are inherently more computationally demanding than matrix operations. Since the code was independently developed without extensive optimization for execution efficiency, we are unable to provide absolute runtime comparisons, and can only offer relative time comparisons.

The code was executed on an Intel Xeon W-2223 CPU with 128 GB of memory, capable of performing computations for the three subregions (119,939×91×3; 174,091×91×3; 207,779×91×3) as well as the combined full-domain reconstruction.

In the manuscript, we have added a corresponding discussion in the second paragraph of the Discussion section.

"However, due to the tensor operations involved, T-DINEOF requires longer computation times to reconstruct the same region. This may limit its application to larger-scale tensors, such as longer time-series images (91 scenes were used in this study) or tensors with more dimensions (three variables in this study)."

We hope the reviewer will understand and appreciate our rationale.

14. Page 22 Lines 23-24: "if the source datasets contain systematic biases, such errors may be propagated". This should be true for any reconstruction method. Please consider adding that tensor methods might be more susceptible to bias propagation because errors in one variable could affect others through the coupled decomposition.

Response: We agree with the reviewer's comment and have incorporated it into the revised Discussion section. Specifically, we added the following sentence: "In particular, tensor-based methods may be more susceptible to such bias propagation, as errors in one variable can influence others through the coupled decomposition."

15. Page 25 Lines 1-2: Is the statement that "the optimal configuration for reconstruction is the SST-SCHL-SSW order" made barely based on the three input orders presented here or all the possible six input orders? If it is the former case, it is inappropriate using "optimal". It should be always cautious using "optimal".

Response: We have revised the corresponding statement to better describe the results shown in Fig. 19. In the revised manuscript, the relevant content reads: "For SST data, the order of input variables has a minor impact, with an average RMSE difference of 0.04 °C. Among the three selected input orders, the configuration that yields the best reconstruction performance is the SST-SCHL-SSW order."