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.

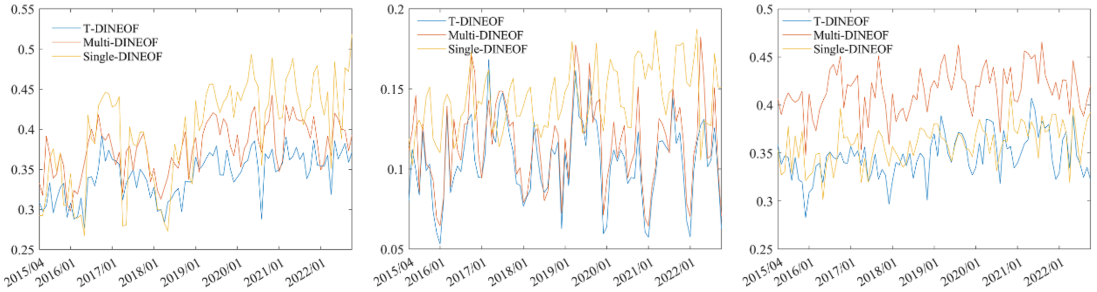
Response: A black ellipse has been added to Fig. 11d.

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.

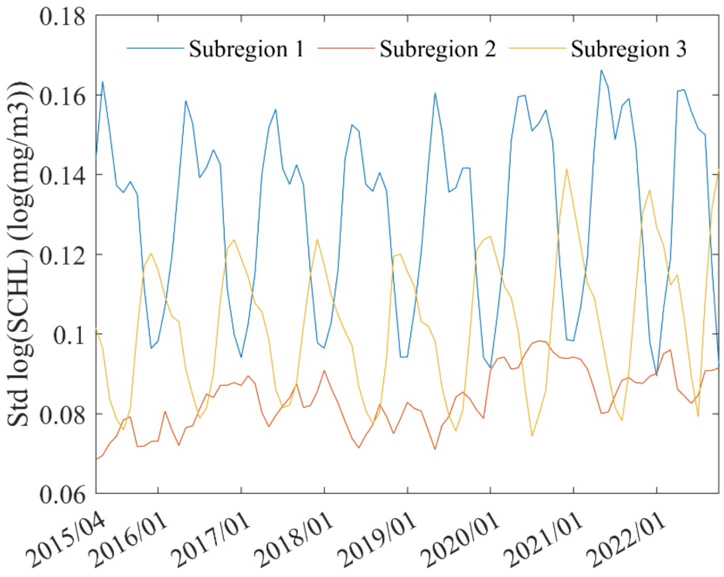


Fig. S2 Temporal std distribution (std) of log(SCHL) in the three study regions.

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.”

To Reviewer 1

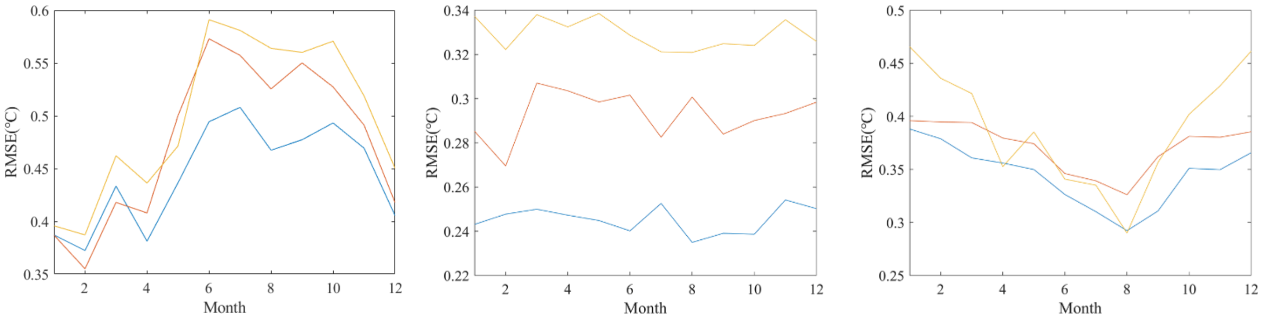
First, we appreciate the reviewer's valuable comments. For your comments, we gave our corresponding explanations and responses below:

1. Examine the total monthly variation (Jan-Dec, not the monthly time series) of RMSE for each variable in each subregion (not over the entire study region) to better demonstrate the inherent seasonal patterns, rather than just some fluctuations during the experimental period.

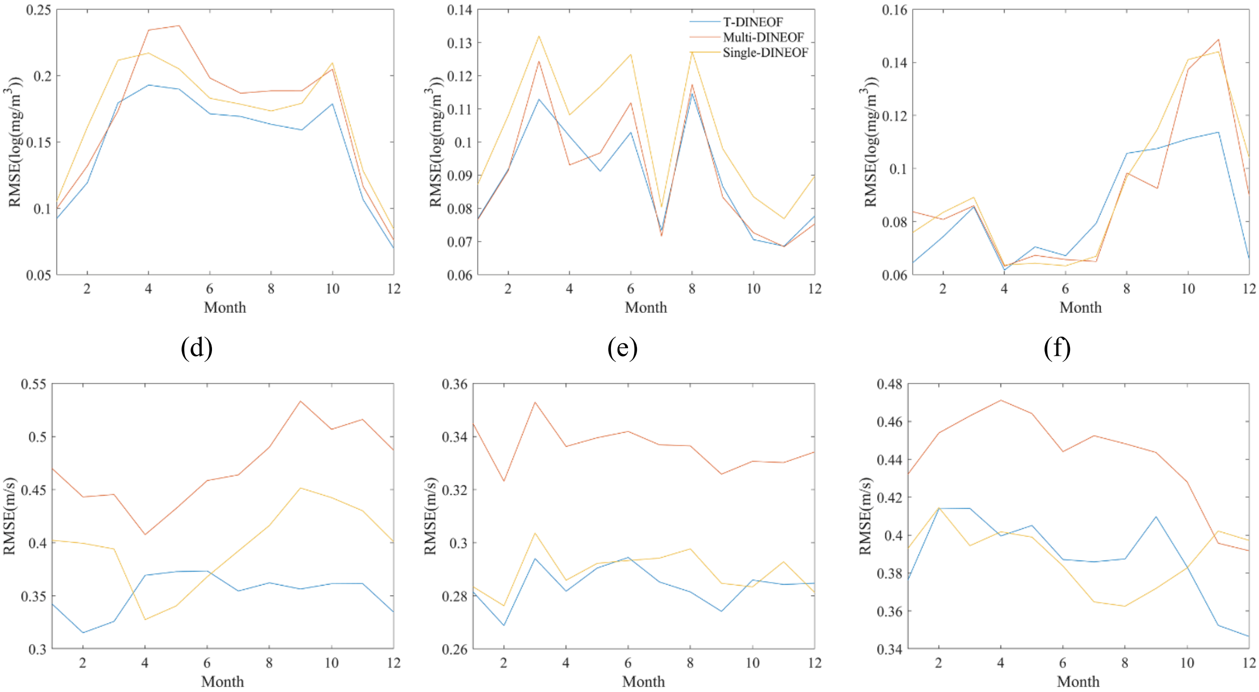
Response: We have regenerated the monthly RMSE distributions of SST, SCHL, and SSW for the three subregions, and accordingly revised the related descriptions in the main text to reflect the monthly RMSE results. Due to space limitations, the monthly RMSE distributions have been included in the Supplementary Material. The revised figures and corresponding descriptions are as follows.

"Meanwhile, the monthly RMSEs of SST, SCHL, and SSW across different subregions were further evaluated (Fig. S3). The results show that, for SST, T-DINEOF achieves the highest reconstruction accuracy in most cases, with particularly notable improvements in subregion 2. Only in subregion 1 during January to March does the reconstruction accuracy of T-DINEOF appear slightly lower than that of Multi-DINEOF. For SCHL, in subregion 1, T-DINEOF achieves the best reconstruction accuracy in most months. In subregion 2, the reconstruction accuracy of T-DINEOF is slightly higher than that of Multi-DINEOF in most months, while Single-

DINEOF yields the lowest accuracy. In subregion 3, T-DINEOF shows higher accuracy from October to the following February, whereas from May to September, its reconstruction accuracy is lower than that of both Multi-DINEOF and Single-DINEOF. For SSW, in subregion 1, T-DINEOF demonstrates higher accuracy in most months except from April to June. In subregion 2, T-DINEOF outperforms both Multi-DINEOF and Single-DINEOF in most months. In subregion 3, T-DINEOF achieves higher reconstruction accuracy from October to the following February and in April, while in the remaining months, Single-DINEOF performs better than T-DINEOF. It is also noteworthy that Multi-DINEOF exhibits the lowest accuracy for SSW. 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)



(d) (e) (f)

(g) (h) (i)

Fig.S3 Monthly RMSEs for (a-c) SST, (d-f) SCHL, and (g-i) SSW in subregions 1-3 (from left to right).

To Reviewer 2

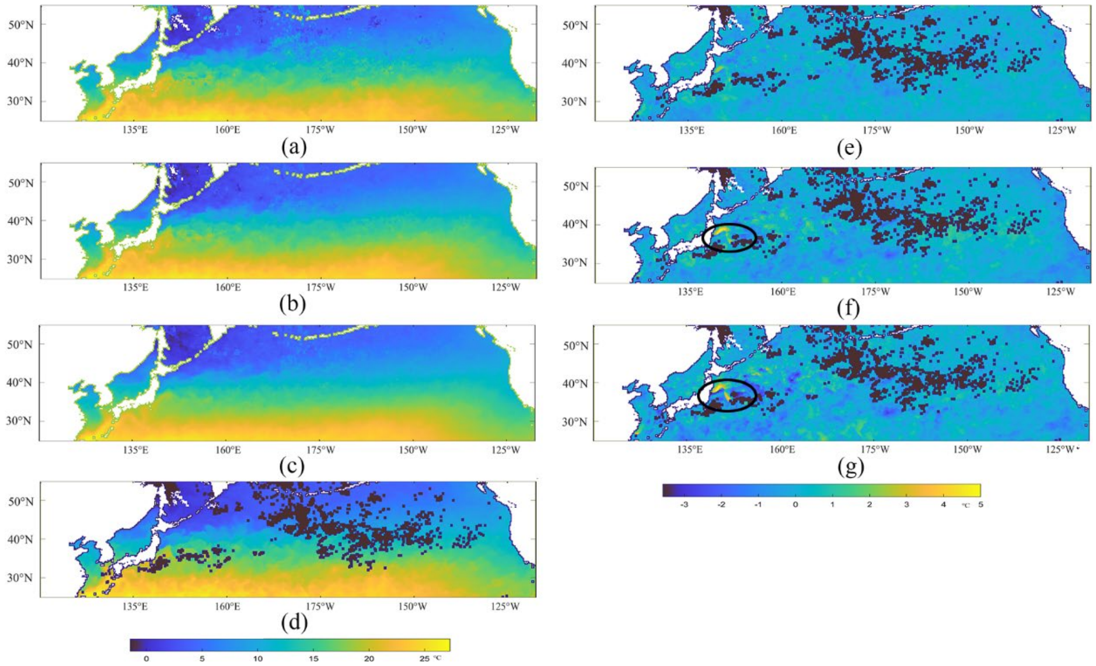
First, we appreciate the reviewer's valuable comments. For your comments, we gave our corresponding explanations and responses below:

1. Figure 7 combines data from three variables (SST, SCHL, and SSW) with different physical units into a single scatter density plot without axis labels or units. The same issue applies to Figure 13. Please either (a) plot each variable separately with proper physical units (°C for SST, mg/m³ for SCHL, and m/s for SSW) on the axes, or (b) if maintaining the combined plot format, clearly state in the figure caption that values are shown in original physical units after inverse normalization, and list the respective units for each variable.

Response: We wish to retain the combined plot format and include the following statement in the figure captions of Figs. 7 and 13: *"*Values are shown in the original physical units after inverse normalization: SST in °C, SCHL in mg/m³, and SSW in m/s.*"*

2. The longitude and latitude labels in Figures 9–11 appear compressed or distorted. Please check.

Response: In the North Pacific region (Figs. 9–11), the longitudinal range is much greater than the latitudinal range, resulting in images of size 120 × 530. To prevent stretching or distortion that could affect the interpretation of the displayed regions, these figures were generated by strictly maintaining the proportional coverage of longitude and latitude, with an aspect ratio of approximately 1:4.42. In the revised version, we regenerated the longitude and latitude labels, as shown below.



3. The validation is performed only against the original satellite-derived data. Given that the original data themselves contain uncertainties and gaps, I suggest validating the reconstructed results against other high-quality monthly products (e.g., reanalysis data) to more rigorously assess the absolute accuracy of T-DINEOF.

Response: First, we believe that reconstructing the target dataset and then evaluating the accuracy between the reconstructed data and the target data at known locations can better reflect the reconstruction performance of the algorithm on the target dataset itself (Both Original DINEOF and Multi-DINEOF are implemented based on this method). If the reconstructed results are compared with other types of datasets, such as reanalysis data, the discrepancies between different data sources may affect the assessment of reconstruction accuracy. Taking SST as an example, reanalysis products integrate multiple SST datasets, and the retrieval algorithms used for these datasets differ from those of the target satellite observations, thereby introducing additional uncertainties. Furthermore, the interpolation procedures used in reanalysis products may generate unrealistic SST values, which can further increase the discrepancies.

To verify this point, we introduced the CMEMS Global Ocean Ensemble Reanalysis product at 1/4° resolution monthly means of temperature and velocity (CMEMS GLOBAL_MULTIYEAR_PHY_ENS_001_031) and the monthly chlorophyll-a product at 1/4° resolution derived from the PISCES biogeochemical model (CMEMS GLOBAL_MULTIYEAR_BGC_001_029) to further evaluate the reconstruction accuracy of T-DINEOF and Multi-DINEOF.

The results are summarized in the following table. Overall, the reconstruction errors relative to the reanalysis and biogeochemical model datasets are larger than those obtained when directly comparing with the target satellite observations. In addition, it can be seen that the errors obtained by the T-DINEOF and Multi-DINEOF methods (first two rows of the table) are comparable to the discrepancies between the satellite observations and the reanalysis data themselves (third row). This indicates that the substantial differences between the target satellite data and the reanalysis products tend to mask the actual reconstruction errors of the algorithms. In other words, the reconstructed results are naturally closer to the target satellite observations, since the reconstruction algorithms are specifically designed to recover the target satellite data.

Therefore, we believe that the relatively large discrepancies between the target observations and the reanalysis datasets may obscure the intrinsic performance evaluation of the reconstruction algorithms. For this reason, we would prefer to retain only the accuracy assessment based on comparisons with the target satellite observations in the manuscript, and we sincerely hope for the reviewer's understanding.

	RMSE	MAE	r	R^2
T-DINEOF	4.8626	3.0245	0.9331	0.7972
Multi-DINEOF	4.8715	3.0288	0.9329	0.7965
Sat-Reanalysis	4.8695	3.0215	0.9328	0.7966

4. To better compare the detail-preserving capabilities of DINEOF, Multi-DINEOF, and T-DINEOF, I suggest adding local standard deviation maps or gradient maps.

Response: We calculated the gradient maps of the SST, SCHL, and SSW data reconstructed by T-DINEOF, Multi-DINEOF, and Single-DINEOF, respectively. The SST gradient maps have been added as Fig. 10 in the main text, while the SCHL and SSW gradient maps have been included in the supplementary file due to space limitations. The three sets of gradient maps are shown below. Overall, compared with the Multi-DINEOF and Single-DINEOF methods, the data reconstructed by the T-DINEOF method preserve richer detail information and exhibit larger gradient magnitudes, particularly in the central and eastern North Pacific, further demonstrating the superior detail-preserving capabilities of T-DINEOF.

We have also added the following description in the main text.

“To analyze the detail-preserving capabilities of Single-DINEOF, Multi-DINEOF, and T-DINEOF, the gradients between adjacent eastward and northward pixels were calculated for the SST, SCHL, and SSW images reconstructed by the three methods. The square root of the sum of squared gradients in the two directions was then used as the pixel gradient magnitude to generate the corresponding gradient maps. The SST gradient maps are shown in Fig. 10, while the gradient maps of SCHL and SSW are presented in Figs. S4 and S5, respectively. Overall, compared with the Multi-DINEOF (a) and Single-DINEOF (b) methods, the T-DINEOF method (c) produces more gradient information with higher gradient magnitudes, indicating a better preservation of fine-scale details. The Multi-DINEOF and Single-DINEOF methods yield relatively low gradient values in the central and eastern North Pacific, failing to adequately capture the detailed structures in these regions. In contrast, although all three methods produce relatively high gradient values in the western North Pacific, the T-DINEOF method preserves substantially richer details, further demonstrating its advantage in detail-preserving capabilities.”

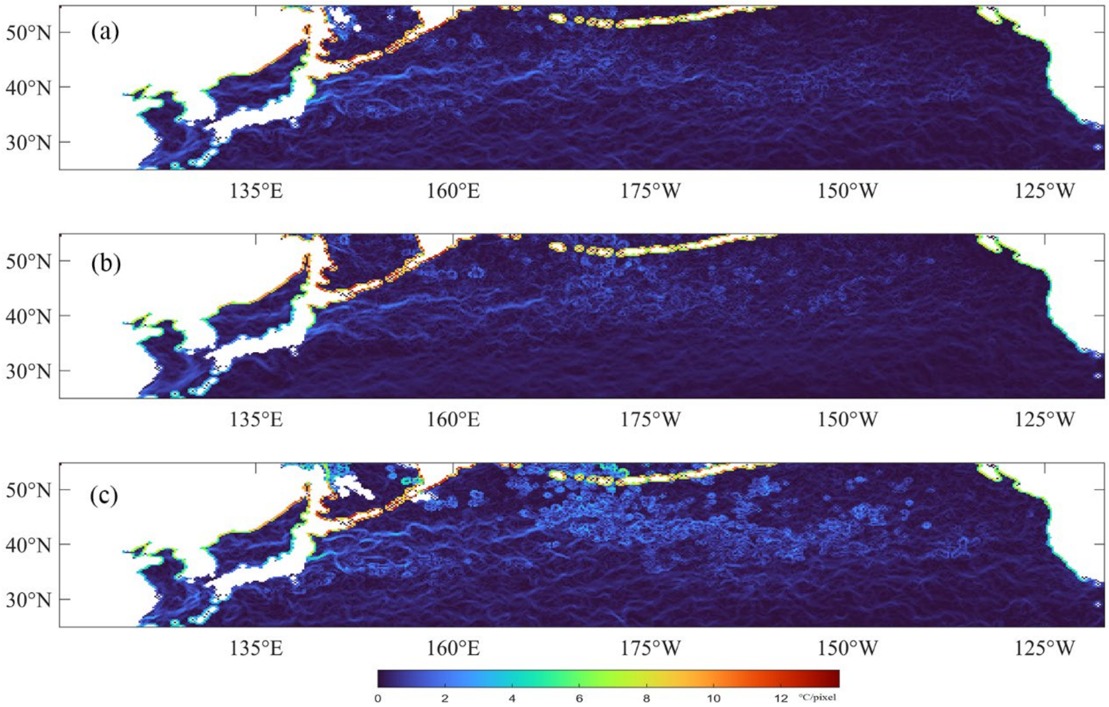


Fig. 1 Gradient maps of SST for April 2022 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

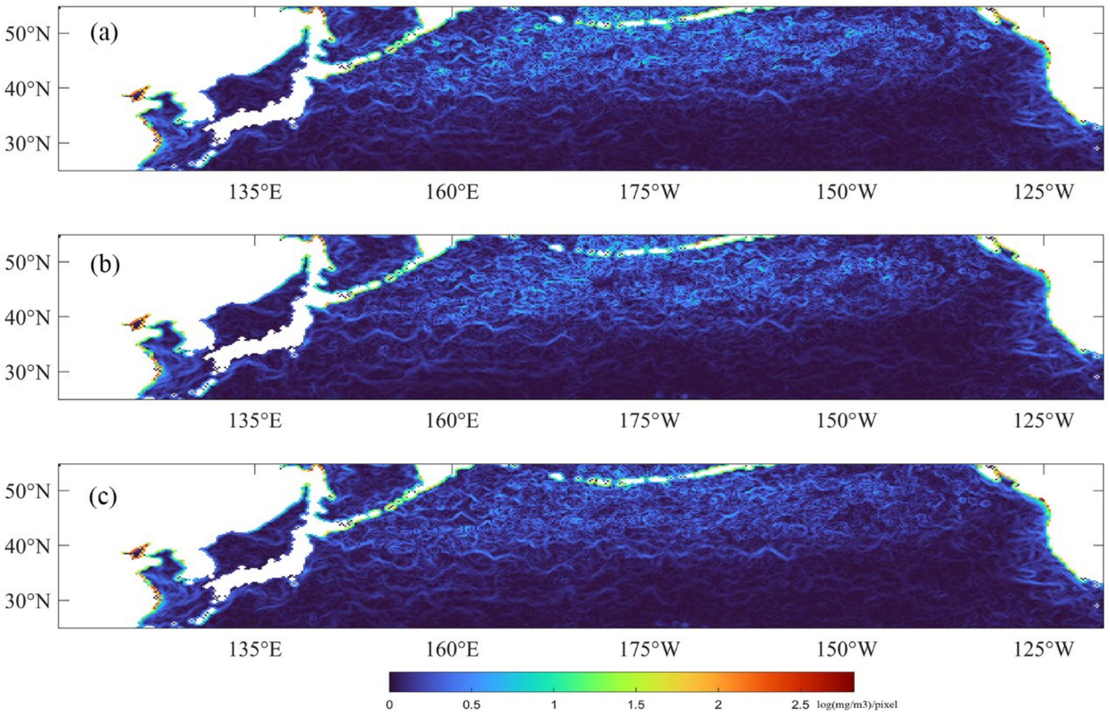


Fig. 2 Gradient maps of SCHL for July 2021 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

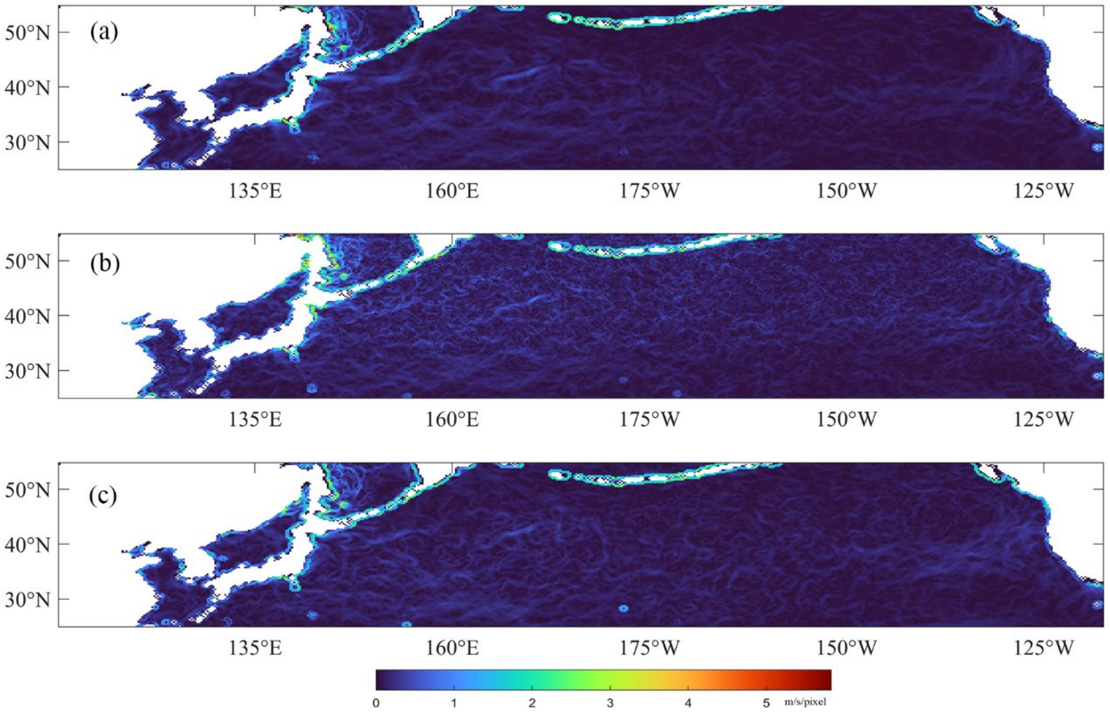


Fig. 3 Gradient maps of SSW for January 2020 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

5. In Figures 7 and 13, a small number of points show underestimation relative to the 1:1 line. What causes this? Please analyze whether these underestimated points correspond to specific geographic regions (e.g., coastal upwelling zones, frontal regions), specific seasons (e.g., summer stratification periods), or specific variables

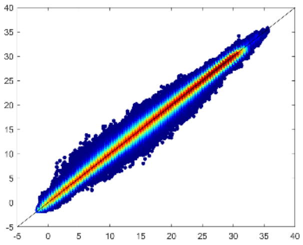
(e.g., high chlorophyll values). This would help clarify the physical mechanisms behind the reconstruction bias.

Response: First, based on the T-DINEOF reconstruction results, we generated separate scatter plots for SST, SCHL, and SSW, as shown in the first row of the figure below. It can be observed that the underestimated points mainly originate from the SCHL data. Based on this observation, we further analyzed the SCHL scatter plots for the three study subregions (second row). The results show that underestimated values mainly occur in the equatorial region (subregion 2) and the Northern Hemisphere mid-latitude region (subregion 1), whereas the SCHL scatter distribution in the Southern Hemisphere mid-latitude region (subregion 3) is comparatively well distributed. This indicates that the underestimated points in Fig. 7 mainly originate from the SCHL data in the Northern Hemisphere mid-latitude and equatorial regions.

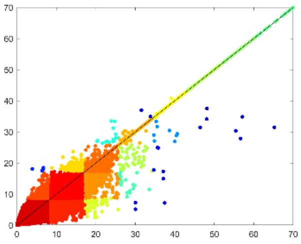
However, due to the relatively limited number of underestimated points, it is difficult for us to further determine whether they are associated with specific geographic regions or particular seasons. Moreover, we believe that a small number of underestimated points alone is insufficient to demonstrate a systematic bias of the reconstruction method for a certain region or season, since most reconstructed values do not exhibit underestimation and only a small fraction show underestimated behavior. We hope for the reviewer’s understanding.

In the revised manuscript, we added the following statement.

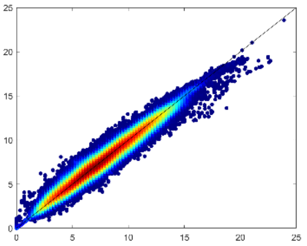
“By analyzing the scatterplot distributions of SST, SCHL, and SSW, it was found that these underestimated values mainly originate from the SCHL data in subregion 1 and subregion 2, indicating that SCHL values may be underestimated in these two subregions.”



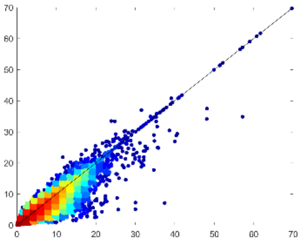
SST



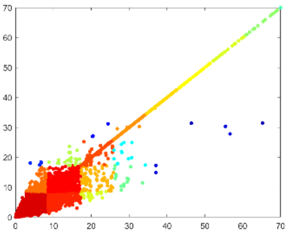
SCHL



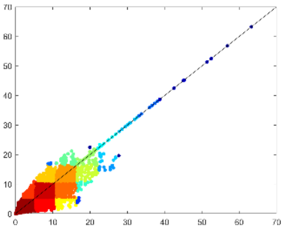
SSW



Subregion 1



Subregion 2



Subregion 3

6. Figure 15 shows obvious periodic fluctuations in SST RMSE over time, particularly in subregion 1. Is this due to lower data availability in summer?

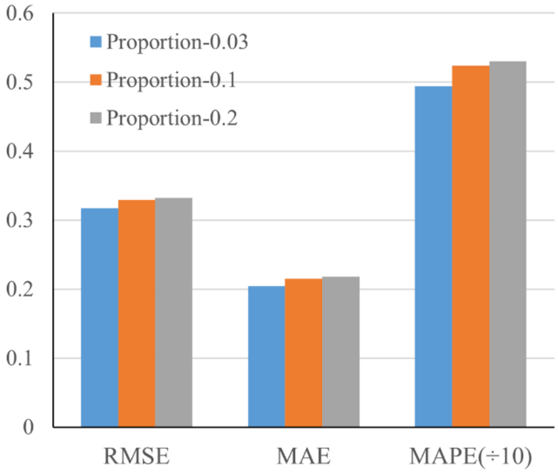
Response: We believe that the periodic variation in SST RMSE is caused by a mechanism similar to that of SCHL, namely the spatial homogeneity of the region.

During summer, SST generally exhibits higher homogeneity, and regions with higher homogeneity tend to produce lower reconstruction errors. In contrast, SST variability is stronger in winter, leading to larger reconstruction errors. Since this explanation has already been discussed in the SCHL section and would be largely repetitive, we chose to retain the discussion only in the SCHL section. We hope for the reviewer's understanding.

7. In Figure 18, the three colored lines (blue, red, grey) are difficult to distinguish. Please revise the figure using more distinct colors, line styles, or symbols to ensure clear visual separation of the three cross-validation proportions.

Response: In the revised version, we replaced the radar chart with a bar chart to better illustrate the differences in reconstruction accuracy. Accordingly, the corresponding text in the main manuscript has been revised as follows.

"Overall, the accuracies obtained under the three proportions are quite similar, while the 3% proportion yields relatively lower error values."



8. On page 10 (line 1), the manuscript states that 3% of existing pixels are selected as cross-validation pixels, while page 23 tests three proportions (3%, 10%, and 20%). Please clarify the rationale for selecting 3% in Section 3.2 (Methodology) rather than deferring this justification to the Discussion section. Readers should understand the basis for this choice when first encountering the method, not later in the paper.

Response: In the revised version, we replaced the corresponding content in Section 3.2 (Methodology) with the following text.

"The same cross-validation pixel selection strategy as in the original DINEOF method was adopted, i.e., 3% of the existing data from the corresponding spatiotemporal matrix of each oceanic variable were randomly selected as cross-validation pixels. In addition, we analyzed the impact of different cross-validation pixel proportions (see Discussion). The results indicate that the choice of proportion has only a minor effect on the reconstruction accuracy, and the 3% setting performs slightly better overall."

To Reviewer 2

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

1. The authors argue that large satellite-reanalysis discrepancies justify omitting independent validation. However, the manuscript mentions daily data experiments (Jan–Mar 2022, Sect. 4.2). Were daily reconstructed results validated against daily reanalysis or buoy data? Monthly averaging may amplify discrepancies; daily comparison would better assess whether the large RMSE reflects true reconstruction error or temporal resolution mismatch. I suggest including daily validation if available.

Response: We still believe that biases among different data sources can affect the accuracy assessment of the T-DINEOF and Multi-DINEOF methods. Therefore, comparing the reconstruction results with their corresponding original data sources can better reflect the actual performance of the algorithms. As discussed in the previous response, compared with the monthly reanalysis data, the reconstruction accuracy of both T-DINEOF and Multi-DINEOF decreased. Because the discrepancies between the original satellite observations and the reanalysis data are relatively large, the resulting errors reflect not only reconstruction uncertainty but also the biases between different datasets.

Similarly, we further evaluated the reconstruction accuracy of T-DINEOF and Multi-DINEOF using the daily average SST at 20 cm depth derived from the C3S global Sea Surface and Sea Ice Temperature Reprocessed product at 0.05° × 0.05° resolution (SST_GLO_SST_L4_REP_OBSERVATIONS_010_024), the 4-km daily chlorophyll-a product derived from the Copernicus-GlobColour processor (OCEANCOLOUR_GLO_BGC_L4_MY_009_104), and the CMEMS Global Ocean Ensemble Reanalysis daily velocity product at 0.25° resolution (GLOBAL_MULTIYEAR_PHY_ENS_001_031). Among these datasets, the SST and SCHL products were resampled to 0.25° resolution to ensure consistency with the SSW data and the reconstructed datasets.

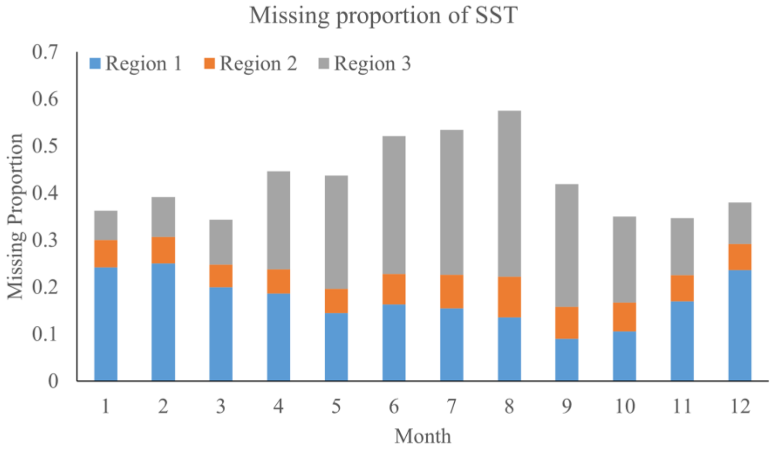
Similar to the results obtained using the monthly datasets, the reconstruction accuracies evaluated with the daily reanalysis data were lower than those evaluated using the satellite observations, and were even lower than the accuracies obtained with the monthly reanalysis datasets. On the one hand, consistent with the conclusions derived from the monthly datasets, the discrepancies between the input satellite observations and the reanalysis products are inherently large. Consequently, the reconstruction errors evaluated against the reanalysis data mainly reflect the differences between data sources rather than the actual reconstruction capability of the methods. Since the reconstruction results of both T-DINEOF and Multi-DINEOF are more strongly constrained by the input satellite observations, the comparison with reanalysis data may not adequately represent the relative reconstruction performance of the two methods.

On the other hand, the daily datasets contain substantially higher missing-data ratios than the monthly datasets. As a result, the reanalysis products are required to estimate a larger number of unknown grid values, which may introduce additional uncertainties and generate more erroneous estimations of oceanic variables. This is likely another reason why the daily datasets exhibit larger errors than the monthly datasets. For this reason, we would prefer to retain only the accuracy assessment based on comparisons with the target satellite observations in the manuscript, and we sincerely hope for the reviewer's understanding.

	RMSE	MAE	r	R^2
T-DINEOF	6.9053	5.2532	0.9002	0.5482
Multi-DINEOF	6.9124	5.2620	0.8994	0.5389
Sat-Reanalysis	6.9058	5.2512	0.9001	0.3248

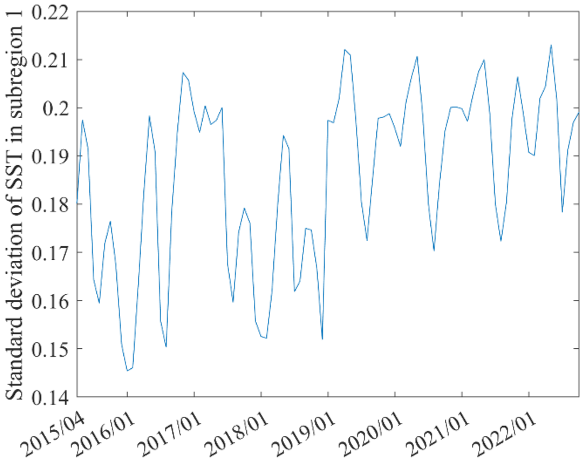
2. I think there may be a misunderstanding in the explanation. The response states that summer SST exhibits "higher homogeneity" leading to "lower reconstruction errors," while winter has "stronger variability" causing "larger reconstruction errors." However, Figure 15a-c clearly shows higher SST RMSE in summer and lower RMSE in winter—the opposite pattern. I would suggest the authors verify whether the seasonal RMSE pattern aligns with missing data proportions.

Response: We found that in subregion 1, summer is generally not the period with the largest amount of missing data, as shown in the figure below. Therefore, similar to the case of SCHL, we believe that variations in the regional parameter itself are the primary cause of the RMSE fluctuations. In our previous response, we stated that the variability is smaller in summer and larger in winter; however, this interpretation appears to be opposite to the actual situation. We would like to apologize to the reviewer for this mistake.



We calculated the monthly standard deviation (std) distribution in subregion 1, as shown in the figure below. The results indicate that the std values are generally higher in summer and lower in winter, suggesting that SST variability is stronger in summer than in winter. In addition, we analyzed the correlation between RMSE and std values, which reached 0.39 in subregion 1. This further supports the conclusion that regional SST variability is a major factor driving the periodic variations in RMSE.

As mentioned in our previous response, in the SCHL section we also investigated the relationship between the periodic variations of RMSE and std. Therefore, to avoid repetitive discussion in the revised manuscript, we did not include a similar analysis for the periodic SST variations. We hope the reviewer can understand this consideration.



To Reviewer 3

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

1. Validation design needs to be described much more clearly.

The manuscript should provide a fully transparent description of how masks were generated, how training and test data were separated, and whether the reconstruction was evaluated in a way that prevents spatial or temporal leakage. For geophysical fields, simple random masking can be misleading because nearby points are correlated, and performance may be overestimated if contiguous structures leak into the training set. The paper should state whether cross-validation was done by space, time, or independent scenes, and whether all methods were tested under identical missingness patterns. Without this, the reported skill gains are difficult to interpret.

Response: The DINEOF, Multi-DINEOF methods and T-DINEOF method proposed in this study are both adaptive reconstruction approaches. Unlike machine learning methods, they do not require training, validation, or test datasets; that is, these methods do not build a reconstruction model, but instead operate directly on the missing matrix/tensor, reconstructing it via matrix/tensor decomposition. For accuracy assessment, the DINEOF, Multi-DINEOF, and T-DINEOF methods all evaluate reconstruction performance by comparing the reconstructed fields with satellite observations at non-missing points. Since the non-missing points in the satellite data are fixed, the accuracy comparison among the three methods is based on the same reference locations. This procedure is also consistent with the standard practice commonly adopted in DINEOF-type methods. Therefore, there is no process of generating masks or splitting data into training and test sets in this study. Currently, DINEOF and its subsequent improved methods all adopt this approach for missing data reconstruction. Furthermore, although 3% of the data were selected as cross-validation points in this study, these points were primarily used to determine the optimal EOF modes rather than to assess reconstruction accuracy. This practice follows the original DINEOF method and its improvements, and the present study adopts the same strategy to determine the optimal modes.

2. The comparison set is likely too narrow for a strong methods paper.

Comparing only against Multi-DINEOF and DINEOF may not be enough to demonstrate that the new method is broadly competitive. Ocean reconstruction and gap-filling literature includes many alternatives, such as EOF-based interpolation variants, low-rank matrix completion, tensor completion approaches, machine-learning-based infilling, and hybrid physical-statistical methods. Even if not all can be included, the paper should justify the chosen baselines and explain why they represent the relevant state of the art. A stronger benchmark suite would make the contribution more convincing.

Response: The DINEOF method is widely used for ocean data reconstruction, and many improved versions have been proposed based on it. The purpose of this study is not to introduce a state-of-the-art ocean reconstruction method, but rather to propose a tensor decomposition-based DINEOF method built upon the existing DINEOF framework, aiming to further improve the reconstruction accuracy for multivariate datasets. We acknowledge that, with the development of artificial intelligence, many machine learning and deep learning methods may achieve higher reconstruction accuracy than DINEOF-type methods. However, DINEOF-type methods still have development potential due to their adaptive nature, which does not require labeled data, the design of an explicit model, or any prior knowledge or assumptions. Therefore, the purpose of this study is to demonstrate that the tensor decomposition-based DINEOF method achieves better accuracy than the current matrix-based Multi-DINEOF reconstruction and provides a new perspective for multivariate reconstruction, rather than to claim that DINEOF-type methods are superior to emerging data-driven or compressive sensing reconstruction approaches.

3. The physical meaning of the reconstructed fields must be demonstrated.

Lower RMSE alone does not prove that the method preserves oceanographically relevant features. The authors should show whether fronts, filaments, coastal gradients, eddy structures, and seasonal patterns remain realistic after reconstruction. For a paper in ocean science, it is important to assess not only pixel-wise accuracy but also whether the reconstructed fields are dynamically plausible and consistent with known ocean variability. Maps, anomaly fields, and regional examples would help substantially.

Response: We calculated the gradient maps of the SST, SCHL, and SSW data reconstructed by T-DINEOF, Multi-DINEOF, and Single-DINEOF, respectively. Typically, oceanographic features such as fronts and eddy structures are derived from gradient fields. The SST gradient maps have been added as Fig. 10 in the main text, while the SCHL and SSW gradient maps have been included in the supplementary file due to space limitations. The three sets of gradient maps are shown below. Overall, compared with the Multi-DINEOF and Single-DINEOF methods, the data reconstructed by the T-DINEOF method preserve richer detail information and exhibit larger gradient magnitudes, particularly in the central and eastern North Pacific, further demonstrating the superior detail-preserving capabilities of T-DINEOF.

We have also added the following description in the main text.

“To analyze the detail-preserving capabilities of Single-DINEOF, Multi-DINEOF, and T-DINEOF, the gradients between adjacent eastward and northward pixels were calculated for the SST, SCHL, and SSW images reconstructed by the three methods. The square root of the sum of squared gradients in the two directions was then used as the pixel gradient magnitude to generate the corresponding gradient maps. The SST gradient maps are shown in Fig. 10, while the gradient maps of SCHL and SSW are presented in Figs. S4 and S5, respectively. Overall, compared with the Multi-DINEOF (a) and Single-DINEOF (b) methods, the T-DINEOF method (c) produces more gradient information with higher gradient magnitudes, indicating a better preservation of fine-scale details. The Multi-DINEOF and Single-DINEOF methods yield relatively low gradient values in the central and eastern North Pacific, failing to adequately capture the detailed structures in these regions. In contrast, although all three methods produce relatively high gradient values in the western North Pacific, the T-DINEOF method preserves substantially richer details, further demonstrating its advantage in detail-preserving capabilities.”

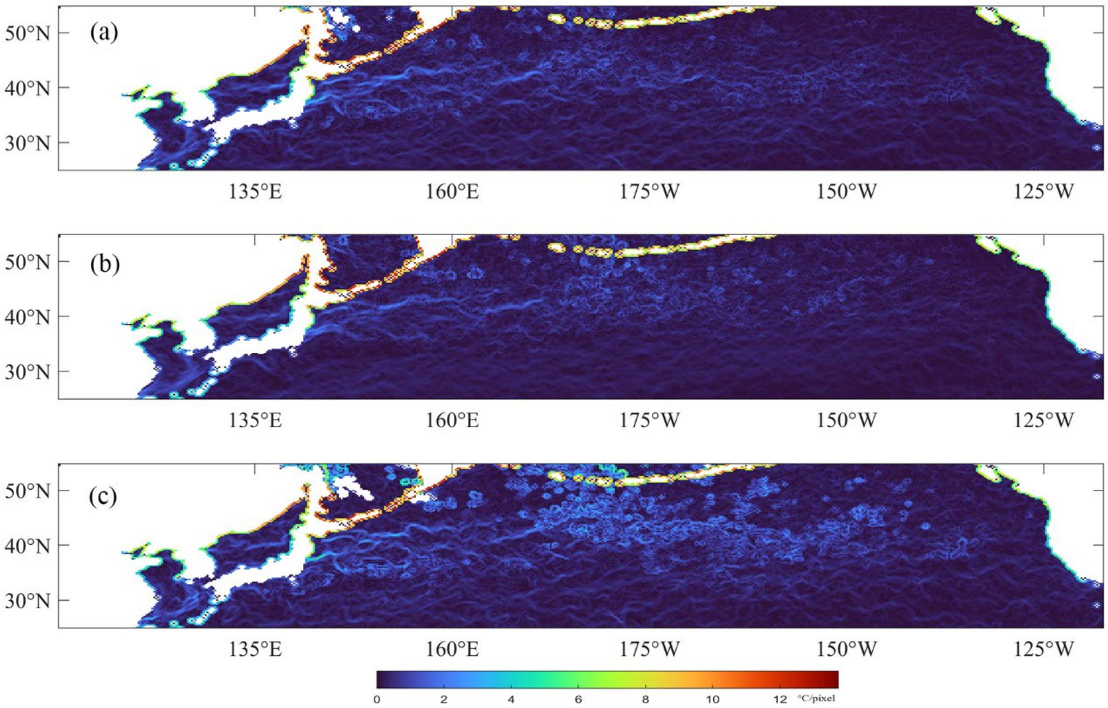


Fig. 1 Gradient maps of SST for April 2022 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

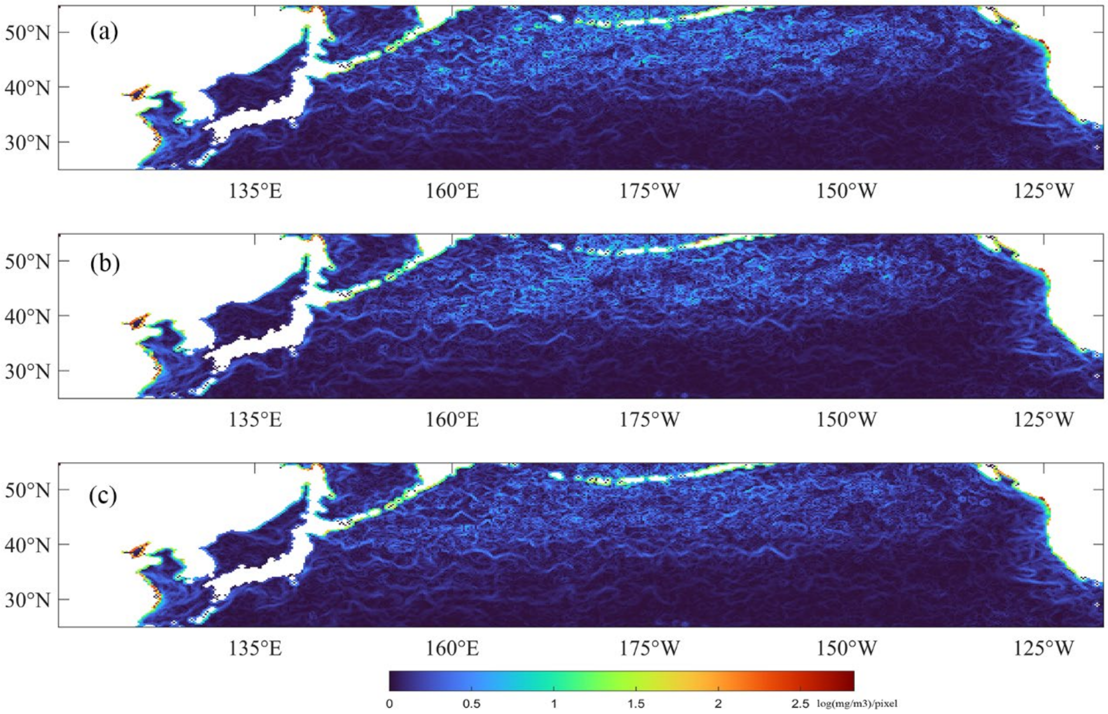


Fig. 2 Gradient maps of SCHL for July 2021 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

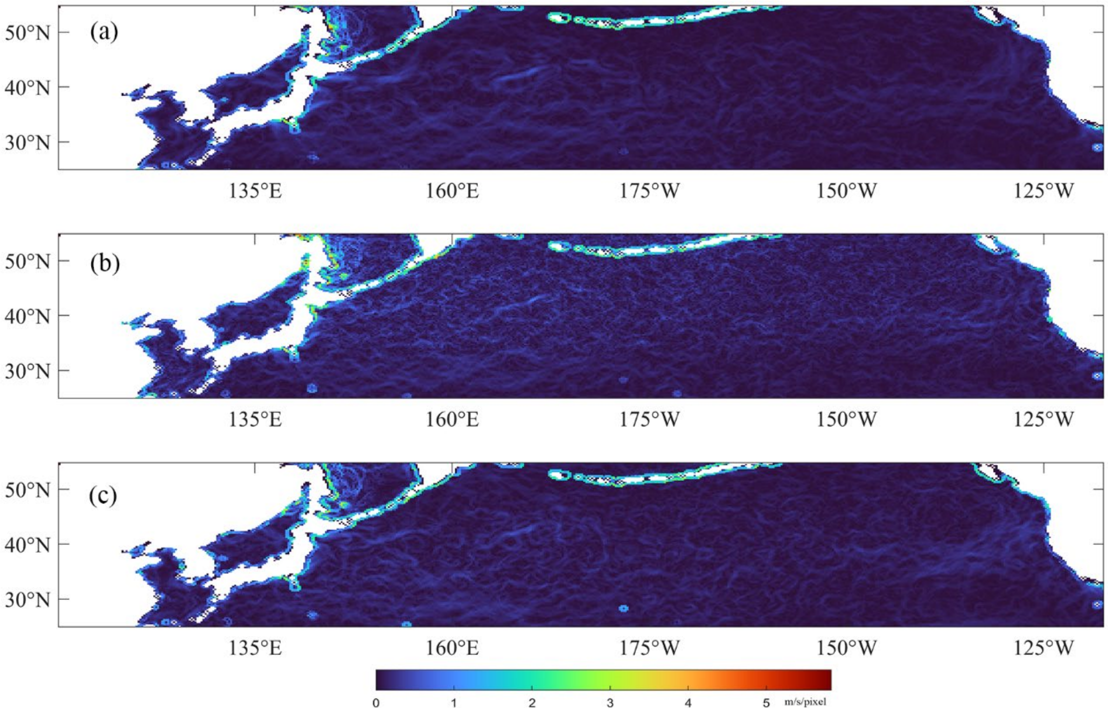


Fig. 3 Gradient maps of SSW for January 2020 in the northern Pacific obtained from (a) Multi-DINEOF, (b) Single-DINEOF, and (c) T-DINEOF.

4. The claim of improved performance in low-correlation regions needs stronger evidence.

The abstract states that T-DINEOF performs better in low-correlation scenarios, which is potentially interesting but also a demanding claim. Low-correlation regions are where multivariate methods can fail if the cross-variable relations are weak or nonstationary. The manuscript should specify how correlation was defined, how these regions were selected, and whether the improvement is statistically significant across multiple cases. If the result depends on a specific dataset or region, the authors should be explicit about that limitation.

Response: In this study, correlations were calculated directly based on the entire input three-dimensional tensor. The correlation coefficients between the input variables SST, SCHL, and SSW were -0.6 for SST–SCHL, 0.03 for SCHL–SSW, and -0.04 for SST–SSW, all of which were statistically significant ($p<0.01$), indicating that SSW has a relatively low correlation with both SST and SCHL. Previous studies have also shown that SST and SCHL are negatively correlated, while both SST and SCHL exhibit low correlation with SSW. Because DINEOF-type methods operate directly on the missing-value matrix/tensor, no specific local regions were selected for analysis, and the correlation calculation was performed over the entire study area. As a result, there are no multiple cases. Reconstruction results demonstrate that, compared with the conventional Multi-DINEOF method, the T-DINEOF method significantly improves the reconstruction accuracy for SSW. Therefore, it can be concluded that T-DINEOF achieves better reconstruction performance for input variables with low correlations.

5. Uncertainty and robustness are not sufficiently addressed.

For a reconstruction method, the paper should discuss sensitivity to missing-data fraction, noise, variable scaling, and tensor rank or hyperparameter choices. The reported improvements may depend strongly on the chosen decomposition rank, normalization strategy, or stopping criterion. A sensitivity analysis would make the method easier to trust and reproduce. At minimum, the paper should include parameter selection details and a robustness check over a range of conditions.

Response: DINEOF is an adaptive reconstruction approach, for which the only key parameter is the determination of the optimal number of modes. In this study, the Method Section provides a detailed description of how the optimal modes are determined, following a procedure similar to that of the original DINEOF and Multi-DINEOF methods. Furthermore, in the Discussion section, we report the specific optimal mode numbers and the corresponding RMSE values for both T-DINEOF and Multi-DINEOF. The main conclusion is that the two methods have similar optimal modes, indicating comparable convergence rates, while T-DINEOF achieves better reconstruction accuracy.

Regarding missing-data fraction and noise, we applied T-DINEOF to daily datasets with a high proportion of missing values (see Supplementary File S1), where the missing fraction exceeded 78%. Despite the high level of missing data, T-DINEOF still outperformed Multi-DINEOF in reconstruction accuracy. This conclusion further confirms the robustness of the T-DINEOF method and its effectiveness in reconstructing datasets with a high proportion of missing values.

As stated in the manuscript (Section 2, Materials), since the ranges of SST, SCHL, and SSW differ, a max–min normalization method was applied to scale each ocean variable (SCHL data were first log-transformed) in each subregion to the range [0, 1]. The maximum and minimum values of each ocean variable in each subregion are provided in Table 1.

During each decomposition, reconstruction is performed using a fixed number of modes, and the reconstruction accuracy is evaluated. If the accuracy does not meet the requirements, the reconstructed tensor is decomposed again using the same number of modes, and the accuracy is recalculated. This process is repeated until the desired accuracy is achieved ($RMSE < 0.00001$) or a stopping criterion is met (100 iterations, Section 3, Methodology), thereby completing the reconstruction for a given number of modes. The number of modes is then increased, and the process is repeated. The number of modes that yields the highest accuracy is selected as the optimal mode for reconstruction. This procedure is similar to that of the original DINEOF and Multi-DINEOF methods. In the Discussion section, we also compare the optimal mode numbers between T-DINEOF and Multi-DINEOF. Therefore, the tensor decomposition process does not involve a specific design or choice of tensor rank.

In addition, this study also analyzed the impact of different cross-validation fractions on reconstruction accuracy (see Fig. 18 in the Discussion). The results confirm that varying the fraction of cross-validation points does not significantly affect the reconstruction accuracy, and the optimal number of modes remains similar. The 3% fraction chosen in this study—also used in the original DINEOF and its improved versions—yields slightly higher accuracy compared to other fractions. Therefore, we retained the 3% cross-validation fraction in this study.

Table 1 Maximum and minimum values of SST, SCHL and SSW in the three subregions.

	SST (°C)		SCHL (log(mg/m^3))		SSW(m/s)	
	Max	Min	Max	Min	Max	Min
Subregion 1	35.65	-1.50	4.05	-4.16	27.05	0
Subregion 2	33.57	0	4.18	-4.56	15.26	0
Subregion 3	31.83	-1.81	3.32	-5.16	24.91	0

6. The treatment of variables with different units and distributions requires careful justification. SST, chlorophyll-a, and wind speed have very different magnitude ranges, statistical distributions, and physical drivers. The manuscript should explain how these variables were normalized and whether the reconstruction operates on standardized anomalies, log-transformed values, or raw

fields. If the method jointly reconstructs variables on different scales without careful preprocessing, one variable may dominate the tensor decomposition. This is an important methodological detail that should be clarified.

Response: As mentioned in response to the previous comment, the normalization methods for different variables are described in Section 2, Materials. Specifically, a max–min normalization was applied to each ocean variable in each subregion. For SCHL, the data were first log-transformed and then normalized using the max–min method, while SST and SSW were directly normalized without transformation. The maximum and minimum values for each variable in each subregion are provided in Table 1. The reconstruction process is based on these normalized data. The original manuscript states: "Since the ranges of SST, SCHL, and SSW differ, a max–min normalization method was applied to scale each ocean variable in each subregion to the range [0, 1]. The maximum and minimum values of each ocean variable in each subregion are provided in Table 1."

7. Computational cost and scalability should be reported.

Tensor methods can be more expensive than matrix-based approaches, especially for large spatial grids or long time series. The paper should provide runtime, memory demand, convergence behavior, and the practical limits of the method. This matters for operational oceanography and large satellite archives. If the method is meant as a practical alternative, the computational overhead must be discussed honestly.

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. As shown in Figures 17 and 18 in the Discussion, the optimal number of modes is typically between 30 and 40, indicating that both the T-DINEOF and Multi-DINEOF methods can reach convergence relatively quickly. 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.

This issue has already been addressed in the manuscript.

"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)."

8. The manuscript should better explain novelty relative to Multi-DINEOF.

The paper should clearly state what is genuinely new in T-DINEOF beyond "tensorizing" a known matrix method. If the novelty is mainly mathematical, the authors should articulate why the tensor formulation changes the reconstruction behavior. If the novelty is algorithmic, they should describe the specific improvement in a way that readers can reproduce. At present, the contribution may read as a natural extension rather than a fully established methodological advance.

Response: We consider the main innovation of this study to be the extension of the widely used matrix-based DINEOF method to a tensor-based formulation. In current oceanographic studies, DINEOF-type methods are commonly employed to reconstruct missing data, making reconstruction accuracy crucial for subsequent data applications. Nearly 20 years since its introduction, the DINEOF method remains fundamentally matrix-based. Whether this framework can be improved or enhanced is the key question that our study aims to address. By extending DINEOF to the tensor domain, this work not only improves reconstruction accuracy but also provides a more reliable data foundation for oceanographic research.

As illustrated in Fig. 1, the Multi-DINEOF framework essentially remains a matrix-based decomposition method. It flattens multiple variables into a concatenated two-dimensional matrix, which inevitably obscures the intrinsic coupling relationships among variables, since different parameters (e.g., SST, SCHL, and SSW) are forced to be represented along the same dimension. As a result, the extracted principal components may mix heterogeneous signals, making it difficult to identify the true inter-variable relationships and thus reducing the physical interpretability of the results. Moreover, matrix-based approaches can only characterize pairwise (i.e., second-order) correlations, making them insufficient for capturing higher-order interactions across spatial, temporal, and variable dimensions.

In contrast, tensor decomposition represents data as a higher-order array (e.g.,space×time×variable), thereby preserving the native multi-dimensional structure of the dataset. This framework allows simultaneous extraction of spatial, temporal, and inter-variable modes, maintaining the inherent coupling among dimensions. Tensor-based models, such as T-SVD decompositions, can reconstruct missing data more accurately by leveraging multi-dimensional constraints, leading to more robust and physically consistent results. Furthermore, each decomposed mode corresponds to a specific physical meaning (e.g., spatial pattern, temporal variation, or variable contribution), enhancing the interpretability of the analysis. Therefore, tensor decomposition provides a more natural and efficient framework for representing and reconstructing complex, multi-variable oceanic or environmental systems.

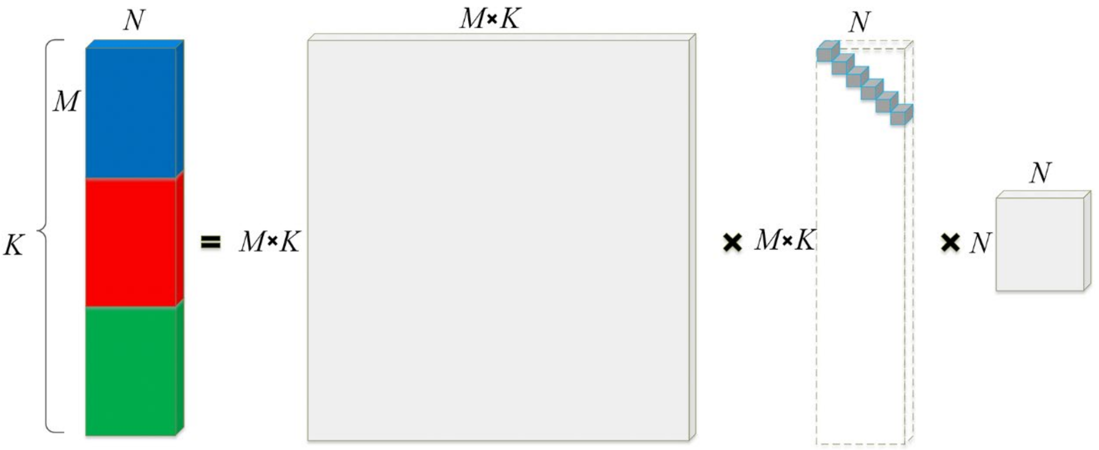


Fig. 1 Matrix decomposition in the Multi-DINEOF method.

Minor comments

9. The abstract is concise, but the manuscript should ensure that the main novelty is stated in the introduction in a way that is understandable to readers outside the immediate DINEOF community.

Response: In the Introduction, we have described the limitations of the DINEOF-type methods and presented the main ideas and innovations of the proposed T-DINEOF approach using both figures and text. Considering the page limit (currently 31 pages, 10,704 words), we did not provide further details on DINEOF-type methods, as this is not intended to be a review article. Nevertheless, we believe that in the current version, the main innovations of the study have been clearly presented in the Introduction.

10. The terminology for the variables should be defined consistently throughout the paper, especially if abbreviations are used for SST, SCHL, and SSW.

Response: We have carefully checked the terminology used throughout the manuscript and did not find any inconsistencies. If any discrepancies exist, they may have been an oversight on our part. We would appreciate it if the reviewer could kindly point out any specific instances of inconsistent terminology so that we can make the necessary corrections.

11. The paper should include more visual comparison figures, ideally with original, corrupted, reconstructed, and error maps shown side by side.

Response: We selected SST data from April 2022, SCHL data from July 2021, and SSW data from January 2020 for visual comparison (Figures 9, 11, and 12). Each of these three figures includes the original, reconstructed, and error maps, presented side by side. As mentioned previously, since DINEOF-type methods do not require generating a mask, there is no corresponding corrupted map; in this case, the original map itself serves as the “corrupted” map containing missing values. Due to space limitations, it is not feasible to show all 91 scenes. In the revised version, we have also included the reconstructed gradient maps to further enhance the visual comparison figures.

12. Metric definitions should be stated explicitly, including any averaging over space, time, or multiple scenes.

Response: We have added the relevant explanations in the Metric Definitions section of Section 3.3 (Methodology). The specific content is as follows:

“where R_{exist} and O_{exist} signify the reconstructed and original values at the existing pixels within all images in the study area during the target period, respectively, and l_{exist} indicates the number of existing pixels.”

13. If the method is applied to satellite products, the preprocessing chain, cloud masking, and regridding procedure should be documented carefully.

Response: We used remote sensing data products distributed by Ocean Color and Remote Sensing Systems, rather than direct satellite observations. The received data products have already undergone preprocessing procedures such as cloud masking and resampling. On one hand, these procedures are well documented and can be directly accessed on the respective websites; on the other hand, they are not the focus of this study and are only marginally related to our research. Considering the page limit, we did not include further details on these preprocessing steps. However, the principles behind the generation of the SST, SCHL, and SSW data relevant to this study have been described in the main text.

14. The discussion should note the main limitations of the approach rather than only emphasizing improvement.

Response: In the second paragraph of the Discussion, we have provided a detailed description of the limitations of the proposed T-DINEOF method, including the use of a single optimal mode, the lack of consideration for temporal correlations, and the computational cost. The specific content is as follows:

“Although T-DINEOF outperforms the original DINEOF and Multi-DINEOF methods, it still relies on a single global optimal mode, which may not be ideal for capturing heterogeneous local dynamics within the study area or achieving the best reconstruction in specific regions and during intermediate iterations. Therefore, integrating improvements from original DINEOF methods and their variants into third-order tensor reconstruction is necessary. Furthermore, this study uses monthly oceanic variables, so temporal correlations are not considered. If daily or weekly data were used, incorporating temporal correlations among various oceanic variables would be essential for enhancing reconstruction accuracy. Additionally, compared to monthly data, daily or weekly datasets typically exhibit higher proportions of missing values. In this study, we demonstrated that T-DINEOF outperforms both Multi-DINEOF and Single-DINEOF in regions with high missing data proportions, suggesting that it may be more advantageous for reconstructing daily or weekly data (Fig. S2). 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). It is evident that, over the same time span, the volume of daily or weekly data is significantly greater than that of monthly data. Given the substantial data volume inherent in a third-order tensor, optimizing computational processes and accelerating processing speeds are crucial for the effective application of T-DINEOF to high-temporal-resolution datasets. 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. Finally, it should be noted that if the source datasets contain systematic biases, such errors may be propagated through the reconstruction process, potentially affecting the accuracy of the final results. 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.”

Main questions

1. How were the missing-data masks generated, and do they preserve realistic cloud-gap or sampling-gap structure?
2. Were the test regions or time periods completely independent from the data used to train the decomposition?
3. How sensitive is the method to tensor rank, normalization, and initialization?
4. Does the method remain advantageous when variables are weakly correlated or when one variable is much noisier than the others?
5. Can the authors demonstrate that the method preserves mesoscale and coastal features, not just global error statistics?

6. How does T-DINEOF compare in runtime and memory use to Multi-DINEOF and DINEOF?

7. Would the method generalize to other ocean variables or other basins beyond the example shown?

Response: For responses to the above issues, please refer to the explanations provided above.

To Reviewer 3

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

1. Validation methodology remains insufficiently justified

The authors explain that DINEOF-type methods are adaptive reconstruction approaches and therefore do not require conventional training, validation, and testing datasets. While this clarification is helpful, it does not fully address the central concern regarding validation rigor.

The primary issue is not whether machine-learning-style training is used, but whether the reconstruction evaluation framework provides sufficiently robust evidence of generalization and avoids overly optimistic estimates arising from strong spatial and temporal autocorrelation.

The current evaluation still appears to rely primarily on reconstruction performance at observed locations rather than demonstrating reconstruction under realistic missing-data structures. While this may follow common DINEOF practice, adherence to previous practice alone does not necessarily demonstrate methodological robustness.

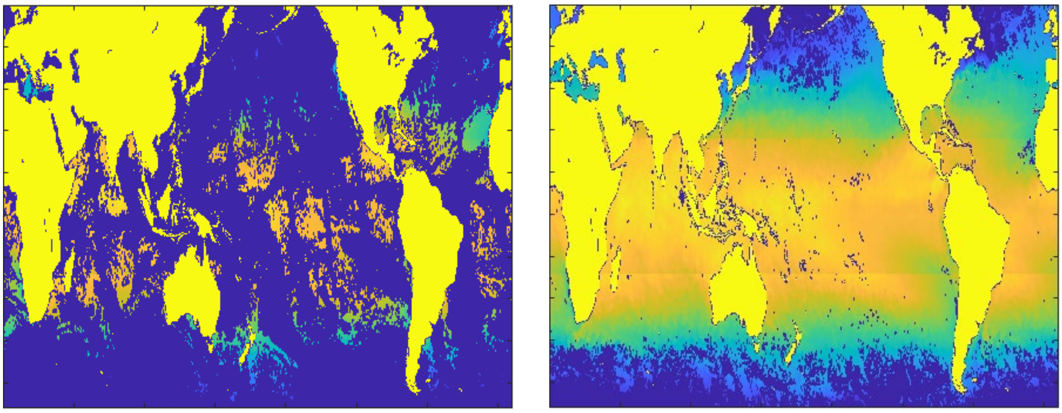
The manuscript would be significantly strengthened through additional validation experiments using more realistic missing-data scenarios, such as:

- contiguous spatial masking,
- withheld scenes or temporal holdouts,
- cloud-like masking structures,
- block-based validation approaches.

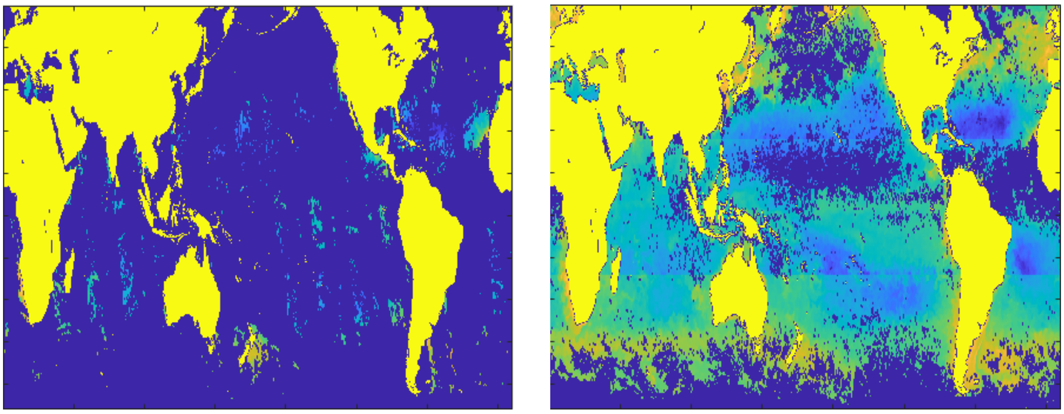
At minimum, the limitations of the current validation framework should be discussed more explicitly.

Response: Following the reviewer's suggestion, we generated a more challenging reconstruction dataset by applying the missing-value masks from the daily SST, SCHL, and SSW observations during 1 January–31 March 2022 to the original monthly datasets. Under this setting, the missing-data ratios increased to 89.88% (SST), 94.82% (SCHL), and 52.29% (SSW) in Subregion 1; 79.94%, 96.36%, and 64.31% in Subregion 2; and 89.28%, 95.76%, and 57.49% in Subregion 3, respectively. Figures below provide examples for April 2015, where the original monthly fields (right panels) are compared with the masked fields (left panels). The substantial increase in missing pixels demonstrates the severity of the imposed missing-data conditions and provides a rigorous test for reconstruction performance. In addition, since each subregion was normalized independently, discontinuities may appear along subregion boundaries.

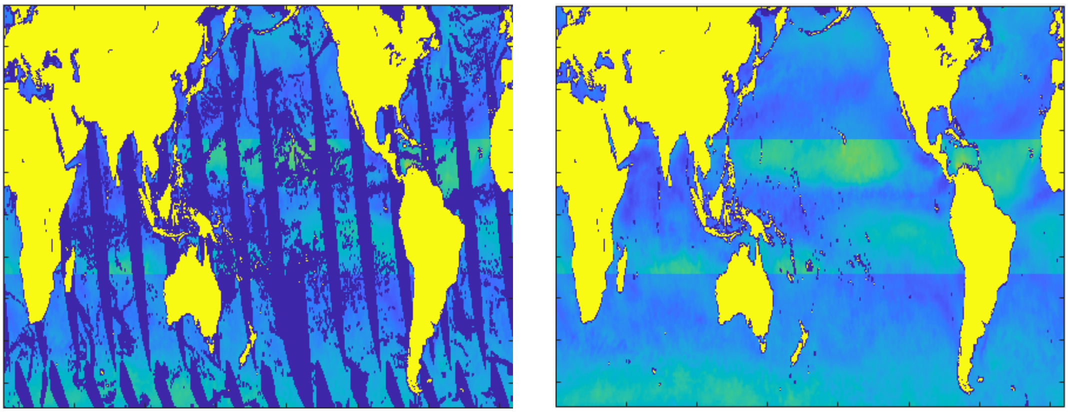
It should be noted that the high proportion of missing data, particularly for SCHL, where the missing rate reaches approximately 95%, poses a substantial challenge for reconstruction. The accuracy is evaluated based on the reconstructed values of daily masked pixels and the corresponding monthly values at the same locations. As expected, this dramatic increase in missing-data ratios results in a pronounced decline in reconstruction performance. Consequently, the accuracies obtained from the masked datasets are significantly lower than those derived from the original datasets. This limitation affects not only DINEOF-type methods but also data-driven approaches, including machine learning and deep learning. We evaluated the reconstruction accuracy of T-DINEOF and Multi-DINEOF methods, as summarized in Table below. The results show that T-DINEOF consistently outperforms Multi-DINEOF, even under these extreme missing-data conditions, demonstrating the robustness of the T-DINEOF approach in handling highly incomplete datasets.



(a) SST with mask (b) original monthly SST at Apr, 2015



(c) SCHL with mask (d) original monthly log(SCHL) at Apr, 2015



(e) SSW with mask (f) original monthly SSW at Apr, 2015

	T-DINEOF		Multi-DINEOF	
	RMSE	Correlation coefficient	RMSE	Correlation coefficient
Subregion1	7.9185	0.8126	7.9624	0.8023
Subregion2	8.0515	0.9738	8.0682	0.9704
Subregion3	9.9562	0.6119	9.9845	0.6072

2. Claims regarding low-correlation performance remain stronger than the evidence currently supports

The authors clarify that correlations were calculated across the entire input tensor and report relatively weak correlations involving SSW. However, the conclusion that T-DINEOF performs better in low-correlation situations still appears stronger than the presented evidence justifies.

The current evidence is primarily based on a single dataset and a single low-correlation variable rather than multiple independent cases or systematically varying correlation regimes.

Therefore, either:

- stronger supporting analyses should be provided, or
- the conclusions regarding low-correlation performance should be moderated.

Currently, the results support improved performance for the presented dataset but do not yet convincingly demonstrate broader behavior under low-correlation conditions.

Response: As this study does not focus on analyzing the correlations among SST, SCHL, and SSW, we have limited the discussion regarding low-correlation scenarios in the conclusions. This ensures that the conclusions accurately reflect the scope of our work while avoiding potential misunderstandings for the reader.

In the revised manuscript, the original content in the Conclusion section has been modified as follows: “Even though the input SSW exhibits low correlation with SST and CHL, the T-DINEOF method is still able to achieve high-accuracy reconstruction of the SSW field.”

3. Physical realism is improved but still only partially demonstrated

The addition of gradient maps represents a meaningful improvement and partially addresses previous concerns regarding physical interpretation.

However, gradient magnitude alone does not fully establish that reconstructed fields preserve oceanographically meaningful structures.

Additional quantitative evidence would strengthen this section, for example:

- variance preservation analyses,

- anomaly structure comparisons,
- feature-preservation metrics,
- spatial spectral analyses,
- evaluation of mesoscale structure retention.

The manuscript has improved substantially in this area, but stronger quantitative evidence would increase confidence that improvements extend beyond pixel-level statistics.

Response: We selected three quantitative metrics to evaluate feature preservation: variance preservation (VP), Structural Similarity Index (SSIM), and anomaly structure (AS). VP was calculated as the ratio of the total variance of the reconstructed field to that of the original field at existing pixels, quantifying the fraction of total variance retained. SSIM was used to assess the structural similarity between reconstructed and original fields, taking into account luminance, contrast, and spatial structure. AS was computed as the correlation between the reconstructed and original anomaly fields, highlighting the ability to preserve fine-scale variations and key spatial features.

For the three study regions, T-DINEOF yielded the following results:

Subregion	VP	SSIM	AS
1	0.9987	0.5626	0.9992
2	0.9996	0.6011	0.9998
3	0.9981	0.5677	0.9991

These quantitative results, together with visual inspection of gradient maps, indicate that T-DINEOF can preserve the dominant spatial structures and major fine-scale features despite some smoothing of local details. The moderate SSIM values likely reflect the intrinsic smoothing behavior of DINEOF-type reconstructions during iterative low-rank approximation.

In the revised manuscript, we additionally included the following analyses and discussions in the revised manuscript:

“In addition to gradient-map comparisons, we introduced three quantitative metrics to evaluate feature preservation: variance preservation (VP), Structural Similarity Index (SSIM), and anomaly structure (AS). VP was calculated as the ratio of the total variance of the reconstructed field to that of the original field at existing pixels, quantifying the fraction of total variance retained. SSIM was used to assess the structural similarity between reconstructed and original fields by considering luminance, contrast, and spatial structure. AS was computed as the correlation between the reconstructed and original anomaly fields, highlighting the ability to preserve fine-scale variations and key spatial features. For the three study subregions, T-DINEOF yielded VP values of 0.9987, 0.9996, and 0.9981, SSIM values of 0.5626, 0.6011, and 0.5677, and AS values of 0.9992, 0.9998, and 0.9991, respectively.

Although the SSIM values (0.56–0.60) indicate that some local-scale differences remain between reconstructed and original fields, the consistently high VP and AS values demonstrate that the dominant variance and anomaly structures are effectively preserved. Combined with the gradient-map comparisons, these results suggest that T-DINEOF retains the major spatial features and fine-scale structures reasonably well, although some degree of smoothing is still present, which is consistent with the intrinsic characteristics of DINEOF-type low-rank reconstructions.”

4. Computational characterization remains limited

The authors acknowledge that T-DINEOF is computationally more expensive and discuss hardware specifications and relative convergence behavior.

However, the manuscript still lacks quantitative characterization of computational performance.

For a methodological contribution introducing a more computationally demanding tensor framework, readers would benefit from:

- approximate runtime comparisons,
- memory usage,
- scaling behavior,
- practical computational limitations.

Even approximate benchmarks would substantially improve the practical relevance of the study.

Response: We appreciate the reviewer's suggestion regarding computational characterization. In the revised manuscript, we recalculated and reported the approximate computational time of the proposed T-DINEOF and Multi-DINEOF methods based on data generated from the daily mask. For subregions 1–3, the running times of T-DINEOF were 29.27h, 35.83h, and 40.68h, respectively, whereas those of Multi-DINEOF were 17.44h, 16.01h, and 13.41h, respectively. Since the total amount of data processed is the same, with the only difference being its organization as a tensor in T-DINEOF and as a matrix in Multi-DINEOF, both methods required similar memory, approximately 6GB per subregion. However, we would like to clarify several important points regarding the interpretation of these benchmarks.

First, the primary objective of this study is methodological development rather than software engineering or computational optimization. The current implementation should therefore be considered a research-oriented prototype intended to demonstrate the feasibility and reconstruction capability of the proposed tensor-based extension. Considerable room for computational optimization likely remains, including memory management, parallelization strategies, and code-level acceleration.

Second, during the reconstruction experiments, multiple regions and/or different algorithms (e.g., T-DINEOF and Multi-DINEOF) were often executed simultaneously to reduce the overall experimental time. Furthermore, the experiments were performed on a single workstation that was simultaneously used for routine research and office work. Consequently, the reported computational costs cannot be interpreted as strict isolated benchmarks for a single region or a single algorithm, and the measured runtime and memory usage should therefore be regarded as approximate rather than rigorous performance metrics.

Third, similar to other DINEOF-type methods, the computational burden increases progressively as the number of retained modes increases because more information participates in the iterative reconstruction process. In the present study, both T-DINEOF and Multi-DINEOF were configured to compute up to 100 modes before selecting the optimal solution, despite the optimal mode number typically occurring around 30–40 modes. This choice was made to prioritize reconstruction capability evaluation and methodological consistency rather than computational efficiency. In practice, strategies such as early-stop criteria could substantially reduce runtime and memory consumption and represent an important direction for future optimization.

Overall, while the computational benchmarks reported in this study provide a preliminary indication of practical cost, they should not be interpreted as fully optimized performance estimates. Our current focus is primarily on extending the DINEOF-type framework and providing new methodological perspectives for multivariate ocean reconstruction. Further engineering-level optimization and software-oriented implementation could be pursued in future work, potentially in collaboration with computational specialists. Therefore, we did not include the exact computational times in the revised manuscript, in order to avoid potentially misleading the readers.