# Response to Reviewer #1

## General Comments

This study presents a global three-dimensional, over 20-year monthly dataset of chlorophyll-a (Chl-*a*) constructed using a multimodal Profile-Surface Transformer (PST) framework. This framework integrates BGC-Argo Chl-*a* profiles, Core-Argo temperature-salinity profiles, and MODIS satellite-derived surface Chl-*a* data. The authors examine the seasonal variability and long-term vertical structure of Chl-*a*, including the depth of the subsurface chlorophyll maximum (SCM). Chl-*a* is a critical variable for characterizing marine phytoplankton biomass and understanding the variability of upper-ocean biogeochemical cycles. Although machine learning/deep learning methods have been widely employed to reconstruct surface Chl-*a* on a global scale, vertical reconstruction remains underdeveloped. The study is novel and significant for understanding the response of the global marine ecosystem and the carbon/nitrogen cycle to the climate change. Overall, the paper is clearly written and easy to understand. However, I have some concerns about the organization, description and analysis of the reconstructed data, which further refinement is needed to meet the publish standard of ESSD. Therefore, I recommended a major revision.

We sincerely thank the reviewer for the careful evaluation of our manuscript and for the constructive comments. These comments were very helpful for improving the organization, methodological clarity, figure presentation, uncertainty discussion, and interpretation of the reconstructed dataset. We have revised the manuscript substantially. We have made a comprehensive revision, include streamlining the Introduction and Discussion, reorganizing the methodology section, adding data-distribution information in the Supplement, clarifying the interpolation strategy and product resolution, revising Figs. 5 and 6, moderating the interpretation of the statistically non-significant global mean SCM-depth trend, adding comparison with satellite ocean-color products and CMIP-class GOBM outputs, and improving the description of validation and uncertainty. Please find the detailed point-by-point responses below.

## Main concerns

**1. Comment:** The current introduction and discussion should be better streamlined

The current introduction is lengthy and contains redundant elements, which could be streamlined to improve readability. I believe that certain sections should be relocated to the discussion. Both the introduction and discussion would benefit from a more thorough literature review. The purpose of the literature review in the introduction is to inform readers about what has been investigated and what remains unexplored, thereby establishing new questions worth addressing. In contrast, the literature review in the discussion serves to compare the current study with existing research and to highlight the novelty of the work. The primary novelty of this research lies in its global reconstruction of three-dimensional chlorophyll-a (Chl-*a*). I recommend incorporating a detailed literature review of surface Chl-*a* reconstruction and Global Ocean Biogeochemical Models (GOBMs) in the introduction. Conversely, the literature review pertaining to regional three-dimensional reconstruction should be expanded in the discussion section (lines 126-146). Additionally, the methodology section (3.1 Problem Definition) should be integrated into the introduction and should not be restated. I had more detailed suggestions on how to reorganize in the specific comments.

**Response:**

Thank you for the great question and the Introduction has been substantially reorganized and streamlined accordingly. In the revised manuscript, the Introduction first introduces the scientific importance of Chl-*a* and its vertical structure, then discusses the limitations of satellite surface Chl-*a* products, the complementarity between BGC-Argo and Core-Argo, and the different limitations of surface-oriented reconstruction methods and GOBMs. Specifically, surface-oriented reconstruction methods remain largely limited to two-dimensional surface fields and cannot fully resolve Chl-*a* vertical structure, whereas GOBMs can simulate 3D Chl-*a* fields but remain affected by uncertainties in ecological parameterizations, process formulations, and model-data mismatches. The detailed discussion of previous regional 3D reconstruction studies has been shortened in the Introduction and revisited in the Discussion when comparing the contribution of the present dataset with previous regional studies. We have also revised the Discussion to better compare the present dataset with previous regional 3D reconstruction studies, satellite ocean-color products, and CMIP-class GOBM outputs. In addition, Section 3.1 has been revised from "Problem definition and overall framework" to "Overall reconstruction framework," and the repeated problem-definition statement has been removed. The revised Section 3.1 now focuses on the overall workflow of model training, inference, and gridded product generation.

**2. Comment:** About the application of the dataset

The authors indicate in several sections that they utilize the data to analyze long-term chlorophyll-a (Chl-*a*) trends and the depth of the subsurface chlorophyll maximum (SCM) over the studied period. However, the statistical results show that the long-term trend is insignificant on a global scale. Therefore, the emphasis on long-term analysis should be moderated. The interannual variability presented in Figure 6b appears significant and warrants more explicit discussion. I recommend that the author team highlight the dataset's applicability for multi-scale analyses, including seasonal, interannual, and long-term trends, rather than focusing excessively on statistically insignificant results.

**Response:**

Thanks for the suggestion. We have revised the manuscript to moderate the emphasis on the statistically non-significant global mean trend and to highlight the multi-scale applicability of the dataset. In the revised Section 5.5, we separate the interpretation of Fig. 6 into regional SCM-depth trends, the global annual mean SCM-depth time series, and the latitude-year distribution. The revised text now emphasizes regional heterogeneity and interannual fluctuations, while stating that the global mean trend is weak and statistically non-significant.

We have also revised the Discussion and Conclusions accordingly. The dataset is now described as supporting multi-scale analyses, including climatological seasonal structure, regional trends, interannual variability, and large-scale latitudinal organization, rather than being framed mainly around a statistically non-significant global trend.

**3. Comment:** For the reconstructed data validation

I recommend that the authors discuss the reconstructed dataset in comparison with CMIP6/7 Global Ocean Biogeochemical Model (GOBM) products. The comparison does not necessarily need to be a full quantitative validation against multiple model outputs, but it would be useful to discuss broad spatial patterns, similarities, differences, and limitations relative to existing model-based products. The current validation framework already includes multiple validation strategies and independent ship-based observations, but such a comparison discussion would help readers better understand the advantage and disadvantages of this dataset.

**Response:**

We agree and have added a qualitative comparison with satellite ocean-color products and CMIP-class GOBM outputs in the Discussion. The revised text now discusses their broad spatial relevance, similarities, differences, and limitations in representing subsurface Chl-*a* structure. Satellite ocean-color records provide long-term, high-resolution observations of surface Chl-*a*, but they cannot directly resolve SCM depth, SCM intensity, or vertical Chl-*a* gradients. CMIP-class GOBMs can simulate Chl-*a* within physically and biogeochemically consistent frameworks and are expected to represent basin- to global-scale Chl-*a* structures. However, simulated SCM depth, SCM intensity, and vertical Chl-*a* gradients may differ among models because of differences in model resolution, ecological process representation, and biogeochemical parameterizations

The revised Discussion further clarifies that the large-scale SCM pattern reconstructed in this study, characterized by deeper SCMs in tropical and subtropical oceans and shallower SCMs in mid- to high-latitude regions, provides an observation-constrained reference for comparing the major spatial organization of subsurface Chl-*a* structure in model-based products. By integrating BGC-Argo Chl-*a* profiles, Core-Argo hydrographic information, and satellite surface Chl-*a* observations, the present dataset provides a complementary 3D reference constrained by both vertical profiles and surface ocean-color information. This revision helps readers better understand the similarities, differences, and complementary roles of the present product, satellite surface records, and CMIP-class GOBM outputs.

**4. A few other main concerns**

**4.1) Figures display concerns:** Figure 5 and 6 are too small, their colorscales are unclear; Therefore, these figures should be enlarged and the colorbars should be changed. Furthermore, a global horizontal distribution figure and representative reconstructed time series (e.g., depth of the subsurface SCM, maximum Chl-*a* concentrations) could be presented.

**Response:**

We thank the reviewer for this helpful suggestion. We have revised Fig. 5 and Fig. 6 to improve readability. For Fig. 5, we added a global horizontal distribution of climatological SCM depth as the first panel. The revised Fig. 5 now includes: (a) global horizontal distribution of climatological SCM depth, (b) global-ocean mean seasonal cycle of reconstructed Chl-*a* as a function of month and depth, (c) climatological latitude-depth section of reconstructed Chl-*a*, and (d) monthly-latitudinal climatology of SCM depth.

After comparing different color scales, we retained the original color scheme because it provides a consistent and interpretable representation of both Chl-*a* and SCM-depth patterns. Instead of changing the color map, we reorganized Fig. 5 into an enlarged 2 × 2 layout, increased the panel size, and improved the labels and color bars. In addition, a non-uniform depth axis was adopted for the Chl-*a* depth-related panels to expand the upper water column and make the 10–300 m upper-ocean and subsurface structure, including the weak subsurface enhancement around 80–120 m, easier to identify.
Regarding the suggestion to present a representative reconstructed time series, we clarified the temporal information provided by the reconstructed product. Specifically, Fig. 6b presents the time series of the global annual mean SCM depth, while Fig. 6c further shows the latitude-year distribution of zonally averaged annual mean SCM depth. These two panels together provide a temporal perspective on SCM-depth variability during 2005–2025. In contrast, Fig. 5 focuses on the climatological seasonal and spatial characteristics of reconstructed Chl-*a* and SCM depth. We have revised the figure captions and related text to make this distinction clearer.

**4.2) Products resolution concerns:** The authors describe resampling the MODIS surface Chl-*a* data to a 0.25° × 0.25° grid during preprocessing (Section 2.3) to balance large-scale patterns and regional details. However, the final reconstructed three-dimensional product is provided at a 1° × 1° resolution. It is recommended that the authors briefly state why the vertical reconstruction can not reach a higher resolution.

**Response:**

We have clarified this point in Section 2.3. The 0.25° MODIS field is used to extract local surface Chl-*a* information for each Argo profile and does not define the resolution of the final 3D gridded product. The final product resolution is ultimately constrained by the spatial density of Argo profiles, monthly sampling support, and interpolation robustness. Therefore, the final gridded 3D Chl-*a* product was generated at 1° × 1° resolution as a compromise between spatial detail, observational support, and product stability.

**4.3) The analysis should be more detailed at the ocean biomes, superbiomes** (Fay and McKinley, 2014), or individual ocean basins. Nevertheless, the current material is sufficiently detailed for the PST framework as presented. I recommended to have those discussions in future work.

**Response:**

We agree with the reviewer that detailed biome-, superbiome-, or basin-scale analysis would be useful for future applications of the dataset.
In the present manuscript, we have kept the focus on the construction, validation, and global-scale characterization of the 3D Chl-*a* product. We have also revised the Discussion to clarify that the dataset is suitable for basin- to global-scale analyses of spatiotemporal variability, vertical ecological structure, and model evaluation (Fay and McKinley, 2014).
Really appreciate the reviewer for providing the great suggestion and we believe this could be a great direction that is valuable for future work.

## Specific Comments

### 1. Introduction

**Reviewer's comment:** The introduction would benefit from a clearer logical structure. The early paragraphs could better connect the importance of Chl-*a*, the limitation of surface observations, and the need for global 3D reconstruction. The review of surface Chl-*a* reconstruction methods and GOBMs should be organized more clearly, with emphasis on why these approaches cannot fully resolve vertical Chl-*a* structure. Some statements about BGC-Argo/Core-Argo and global-scale extension appear repetitive and could be merged and condensed. The discussion of recent regional 3D reconstruction studies could be shortened in the introduction and revisited in the discussion to better highlight the novelty of the present global dataset.

**Response:**

We have revised the Introduction following this suggestion. The revised Introduction now follows a clearer sequence: the importance of Chl-*a* and vertical structure, the limitations of satellite surface observations, the complementarity and limitations of BGC-Argo and Core-Argo, the different limitations of surface Chl-*a* reconstruction methods and GOBMs, and the remaining need for a globally consistent 3D Chl-*a* product. We clarified that surface-oriented reconstruction methods cannot fully resolve Chl-*a* vertical structure. The GOBMs can represent 3D fields but are affected by uncertainties in ecological parameterizations, process formulations, and model-data mismatches. Repetitive descriptions of BGC-Argo, Core-Argo, and global-scale extension have been merged and condensed. The description of previous regional 3D reconstruction studies has also been shortened in the Introduction and revisited in the Discussion to highlight the novelty of the present global gridded product.

### 2. Data sources

**Reviewer's comment:**2.1 In Section 2.1, a data distribution figure for Bio-Argo and Core Argo should be placed in the Supporting Information and cited in both Sections 2.1 and 2.2. This would help readers understand the "observationally sparse regions" mentioned later in line 735.

**Response:**

We have added a data-distribution figure to the Supporting Information. In the revised Section 2.1, a statement is added: "*Fig. S1a compares the full-period 1° spatial density of the Core-Argo profiles used for inference with that of the BGC-Argo Chl-a profiles actually used for model training, validation, and testing*." We also state that Fig. S1b compares the annual profile counts of the Core-Argo inference profiles and the used BGC-Argo profiles on a logarithmic scale. In Section 2.2, we now explicitly state that the spatial and temporal distributions of the Core-Argo profiles are summarized together with the used BGC-Argo profiles in Fig. S1. This comparison highlights the much broader spatial and temporal coverage of Core-Argo relative to BGC-Argo and helps readers identify regions with sparse BGC-Argo observational support, thereby providing context for the interpolation strategy and the uncertainty discussion in later sections.

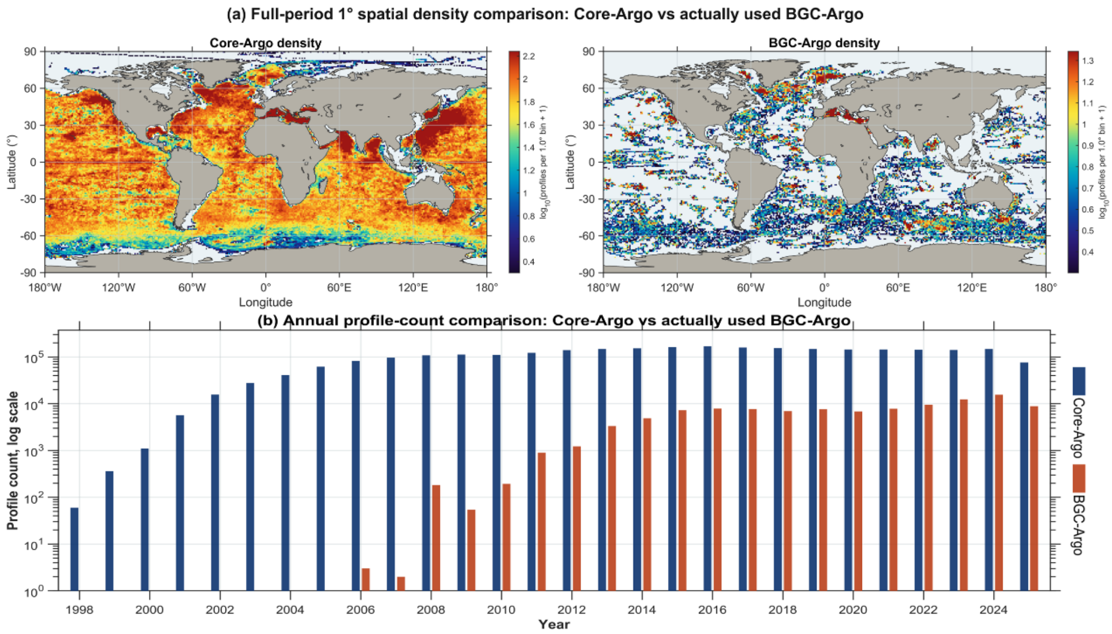


**Figure S1. Distribution comparison of Core-Argo inference profiles and used BGC-Argo Chl-*a* profiles.**

**3. Methodology**

**Reviewer's comment:** 3.1 The problem has already been defined in the introduction. Therefore, there is no need to restate it here. Please remove "Problem definition" from the title and L199-L200.

**Response:**

We have revised Section 3.1 accordingly. The title has been changed to "Overall reconstruction framework," and the repeated problem-definition statement has been removed. The revised section now focuses on the workflow of the reconstruction framework, including the training phase, inference phase, and final gridded product generation.

**Reviewer's comment:** 3.2 L338: Is it reasonable to set the maximum search radius to 700 km? 700 km is an exceptionally large distance that can span multiple dynamical regimes. Applying such a wide radius, especially as a fallback in data-sparse regions (e.g., Southern Ocean), will likely introduce spurious signals and over-smooth the vertical structure of Chl-*a*. The authors should justify why 700 km is optimal compared to smaller alternatives.

**Response:**

Thank you for this important comment. We agree that 700 km is a large spatial scale and may potentially span different dynamical regimes if used as a regular interpolation scale. Therefore, we have clarified in the revised manuscript that 700 km should not be interpreted as the nominal horizontal resolution or the regular interpolation scale of the product. Instead, it is used only as the final fallback radius in the adaptive interpolation procedure. The algorithm always attempts smaller search radii first and selects the smallest radius that provides sufficient neighboring profile support. The 700 km radius is used only when the smaller radii cannot provide enough neighboring profiles. This fallback setting therefore reflects uneven Argo sampling support rather than an assumption of dynamical homogeneity over 700 km.

To quantitatively assess the reasonableness of this setting, we performed a random holdout cross-validation experiment. A subset of inferred profiles was withheld from the interpolation, and the remaining profiles were used to reconstruct Chl-*a* at the withheld profile locations under different maximum-radius configurations. The interpolated values were then compared with the withheld inferred Chl-*a* values. The validation was conducted at individual depth levels of the withheld profiles, so it directly evaluates the depth-resolved interpolation performance and helps assess whether the larger fallback radius degrades the reconstructed vertical Chl-*a* structure.

The results showed that the max700 configuration achieved high interpolation coverage while maintaining low interpolation errors. Specifically, the max700 configuration reached a coverage of 99.91%, with RMSE = 0.02373 mg $m^{-3}$, MAE = 0.00621 mg $m^{-3}$, and R = 0.9822. Compared with smaller maximum-radius configurations, the use of 700 km mainly reduced interpolation gaps in data-sparse regions and did not substantially increase the overall interpolation error.

We also calculated the actual usage frequency of each selected adaptive radius. The 700 km radius accounted for 12.12% of adaptive interpolation cases, indicating that it was not the dominant interpolation radius. As shown in Fig. S2, the contribution of 700 km radius selections was not uniformly distributed over the global ocean, but was mainly concentrated in high-latitude regions, marginal seas, coastal-boundary areas, and other sparsely sampled regions. This spatial pattern further supports the interpretation that 700 km acts as a limited fallback radius in low-support regions rather than imposing broad smoothing over the global product.

We do not claim that 700 km is universally optimal. Rather, we interpret it as a practical upper-bound fallback radius that balances spatial coverage and interpolation error under uneven sampling conditions. We also recognize that grid cells requiring larger radii have weaker local profile support. Therefore, to make this limitation transparent to users, we have added interpolation diagnostic variables to the product, including the selected interpolation radius, the number of neighboring profiles, the adaptive-interpolation failure flag, the nearest-fill flag, and the nearest-fill distance. These variables allow users to identify grid cells with weaker interpolation support and apply stricter filtering if needed.

Overall, the 700 km radius is not used as the nominal resolution or regular interpolation scale of the product, but as a constrained fallback radius for sparsely sampled regions. The random holdout cross-validation, the radius-usage statistics, and spatial distribution of 700 km usage jointly indicate that this setting is reasonable for maintaining spatial continuity while limiting interpolation error.

**Table 1. Random holdout cross-validation results for different maximum adaptive interpolation radii**

A subset of inferred profiles was withheld from the interpolation, and the remaining profiles were used to reconstruct Chl-*a* at the withheld locations. All configurations used the same adaptive interpolation logic and parameters, with only the maximum search radius changed.

| Radius configuration | Maximum radius (km) | Coverage (%) | MAE (mg $m^{-3}$) | RMSE (mg $m^{-3}$) | Bias (mg $m^{-3}$) | R | $R^2$ |
|---|---|---|---|---|---|---|---|
| max100 | 100 | 48.86 | 0.00437 | 0.01899 | -0.000153 | 0.9885 | 0.9772 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| max200 | 200 | 82.25 | 0.00527 | 0.02143 | -0.000138 | 0.9858 | 0.9717 |
| max300 | 300 | 94.98 | 0.00574 | 0.02258 | -0.000136 | 0.9840 | 0.9683 |
| max500 | 500 | 99.57 | 0.00616 | 0.02359 | -0.000102 | 0.9824 | 0.9652 |
| max700 | 700 | 99.91 | 0.00621 | 0.02373 | -0.000089 | 0.9822 | 0.9647 |

Table note: Coverage represents the fraction of withheld Chl-*a* values that could be reconstructed by the adaptive interpolation stage. MAE, RMSE, Bias, R, and $R^2$ were calculated by comparing interpolated Chl-*a* values with the withheld inferred Chl-*a* values. The nearest-fill step was not included in this cross-validation, so the table evaluates the adaptive-radius interpolation itself.

**Table 2. Actual usage of different adaptive interpolation radii**

| Radius | Count | Usage frequency among adaptive interpolation cases |
|---|---|---|
| 100 km | 13,436,450 | 3.06% |
| 200 km | 98,512,700 | 22.41% |
| 300 km | 139,851,450 | 31.82% |
| 500 km | 134,471,200 | 30.59% |
| 700 km | 53,275,600 | 12.12% |
| Total adaptive success | 439,547,400 | 100.00% |

Table note: Radius usage was calculated only for grid cells successfully interpolated by the adaptive-radius stage. Nearest-fill cases were not counted as radius usage.

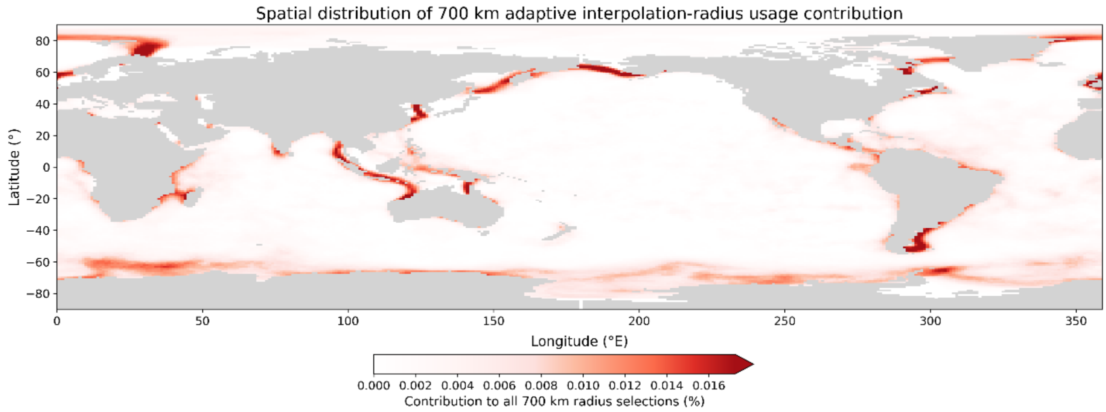


**Figure S2. Spatial distribution of the usage contribution of the selected 700 km adaptive interpolation radius.**

The contribution was calculated as the fraction of all 700 km radius selections contributed by each ocean grid cell during the adaptive interpolation stage. Larger values indicate regions that contributed more frequently to the total use of 700 km fallback radius. Land areas are shaded grey.

**4. Validation and uncertainty assessment**

**Reviewer's comment:**

4.1 L497-L506: should be moved to the Methods section.

**Response:**

We have moved the description of ship-based CTD preprocessing and matching strategy to the Methods section. A new Section 3.6, "Independent ship-based validation data and matching strategy," has been added. This section describes the standardization of ship-based observations, deep-background correction, matching strategy for inferred-profile validation, matching strategy for product-level validation, and vertical averaging within the corresponding depth window. Section 4.2 now refers to Section 3.6 instead of repeating the methodological details.

**5. Examples of spatiotemporal variation characters based on the reconstructed results**

**Reviewer's comment:**

5.1 I suggest provide a global horizontal distribution figure first, this can be the SCM depth/maximum chlorophyll concentrations. I also suggest providing a representative monthly timeseries, if appropriate, and discussing the reconstructed results in comparison with CMIP6/7 GOBM products.

**Response:**

We thank the reviewer for this helpful suggestion. To better demonstrate the ability of the reconstructed 3D Chl-*a* dataset to reproduce known spatiotemporal ecological patterns and to illustrate its applicability for multi-scale analyses, we have expanded Section 5 by adding a new summary figure and corresponding explanatory text.

Specifically, we added a new Figure 5 (Figure S3), entitled "Climatological characteristics of reconstructed Chl-*a* and SCM depth", which summarizes the main climatological features of the reconstructed fields. The figure includes: (a) the global horizontal distribution of climatological SCM depth; (b) the global-ocean mean seasonal cycle of reconstructed Chl-*a* as a function of month and depth; (c) the climatological latitude–depth section of reconstructed Chl-*a*; and (d) the monthly–latitudinal climatology of SCM depth. These additions address the reviewer's suggestion to provide a global horizontal distribution figure and representative seasonal characteristics of the reconstructed results.

We have also revised the Discussion to clarify the relationship between the reconstructed product and CMIP-class GOBM outputs. In the revised Discussion, we state that, compared with satellite ocean-colour products and CMIP-class GOBM outputs, the present dataset provides a complementary, observation-constrained 3D reference for subsurface Chl-*a* structure. We further clarify that satellite ocean-colour records cannot directly resolve SCM depth, SCM intensity, or vertical Chl-*a* gradients, whereas CMIP-class GOBMs may differ in their simulated SCM depth, SCM intensity, and vertical Chl-*a* gradients because of differences in model resolution, ecological process representation, and biogeochemical parameterizations. In this context, the reconstructed large-scale SCM-depth pattern, characterized by deeper SCM depths in tropical and subtropical oceans and shallower SCM depths in mid- to high-latitude regions, provides an observation-constrained reference for comparing the major spatial organization of subsurface Chl-*a* structure in model-based products. Newly added figure caption and representative text are reproduced below.

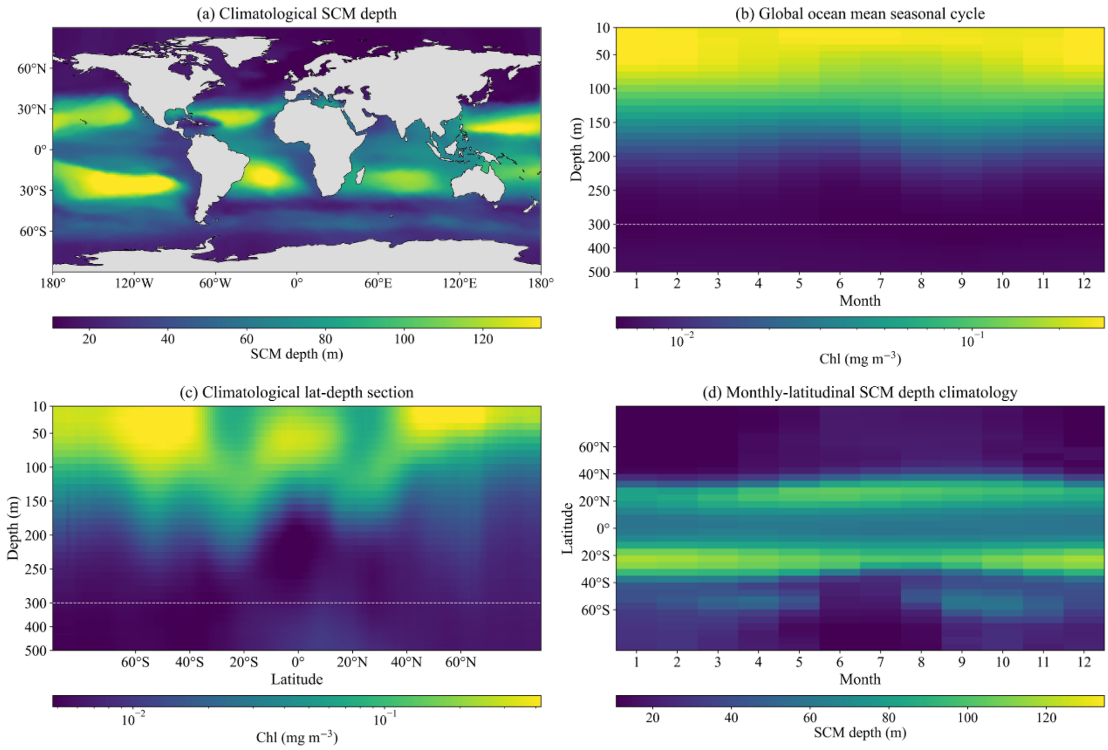


**Figure 5 (Figure S3). Climatological characteristics of reconstructed Chl-*a* and SCM depth. (a) Global distribution of climatological SCM depth. (b) Global-ocean mean seasonal cycle of reconstructed Chl-*a* as a function of month and depth. (c) Climatological latitude–depth section of reconstructed Chl-*a*. (d) Monthly–latitudinal climatology of SCM depth.**

Added text in Section 5:

"*To evaluate the ability of the reconstructed 3D Chl-a dataset to reproduce known patterns of ocean ecological structure and to further demonstrate its applicability for multi-scale analysis, this section examines the mean seasonal and spatial patterns of reconstructed Chl-a and SCM depth, as well as regional trends and interannual variability in SCM depth during the reconstruction period.*"

Added text in Section 5.1:

"*The global horizontal distribution of climatological SCM depth shows pronounced spatial heterogeneity across the reconstructed ocean domain (Fig. 5a). Deeper SCM depths are mainly found in subtropical gyre regions, whereas shallower SCM depths occur more frequently in high-latitude waters, equatorial regions, and productive marginal seas. This pattern is broadly consistent with previous large-scale analyses of Chl-a maximum depth, which reported deeper maxima in tropical and subtropical oceans and shallower maxima in mid- and high-latitude regions (Yasunaka et al., 2022).*"

These revisions make the spatiotemporal characteristics of the reconstructed dataset more explicit and clarify that the dataset can serve as a complementary, observation-constrained reference for evaluating subsurface Chl-*a* structure in CMIP-class GOBMs.

**Reviewer's comment:**

5.2 Figure 5: The panels are too small to be seen clearly. Please reorganize them as upper, middle, and lower panels, and use a more distinct colormap (e.g., jet). Regarding Figure 5c, the Northern and Southern Hemispheres appear asymmetric. Please add a description of this asymmetry and provide a brief explanation.

**Response:**

We have revised Fig. 5 to improve its readability. After comparing different color scales, we retained the original color scheme because it provides a consistent and interpretable representation of both Chl-a and SCM-depth patterns. We reorganized Fig. 5 into an enlarged 2 × 2 layout, increased the panel size, and improved the labels and color bars. For the Chl-a depth-related panels, we adopted a non-uniform depth axis to expand the upper water column and make the 10–300 m upper-ocean and subsurface structure easier to identify.

We also revised the text describing the monthly-latitudinal SCM-depth climatology and the latitude-depth structure. The revised manuscript now more clearly describes the hemispheric asymmetry, including the deeper and more persistent SCM structure in the southern subtropics, and relates this pattern to the large-scale latitudinal organization of SCM depth.

**Reviewer's comment:**

5.3 Section 5.4: Given that the long-term trend is statistically insignificant, this section should not discuss it extensively. Both interannual variability and the long-term trend should be addressed, as Figure 6 includes both: panels (b) and (c) show interannual variability, while panel (a) shows the regional long-term trend. These should be separated into two distinct figures. Climate-mode analyses, such as during typical El Niño/La Niña years, could be mentioned as potential future applications rather than being required in the current manuscript.

Lines 671–672 conclude the main physical controls on SCM depth—particularly light availability, stratification intensity, and nutrient supply—remained broadly stable over 2005–2025. This conclusion may need more cautious wording in relation to the study by Cheng et al. (2025, NREE), which found that the ocean has become more stratified in a warming climate. It may also need to be better reconciled with analyses from data-assimilated global ocean circulation models. The authors should clarify that the absence of a statistically significant global mean SCM-depth trend does not necessarily confirm that the underlying physical controls remained stable over the same period.

**Response:**

We have substantially revised this section. The revised Section 5.5 now discusses regional trends, the global annual mean time series, and the latitude-year structure separately. Fig. 6a is interpreted as showing regionally heterogeneous SCM-depth trends. Fig. 6b is interpreted as showing a weak and statistically non-significant global mean trend, with interannual fluctuations being more prominent than a monotonic global-mean trend over the study period, in consistent with Cheng et al (2025). Fig. 6c is now described as mainly highlighting the persistence of the large-scale latitudinal organization of SCM depth rather than pronounced year-to-year changes.

We considered the reviewer's suggestion to separate the regional trend and interannual variability into two figures. In the revised manuscript, we retained them within a single figure because all three panels describe complementary aspects of SCM-depth variability during the same reconstruction period and are more easily compared when presented together. To avoid ambiguity, we revised the figure title, caption, and Section 5.5 text to clearly distinguish the regional trend map in Fig. 6a from the temporal information shown in Figs. 6b and 6c.

We have also removed the previous statement that the main physical controls on SCM depth remained broadly stable over 2005–2025. In the revised Discussion, we state more cautiously that the coexistence of regionally heterogeneous SCM-depth trends and a persistent latitude-dependent background should be interpreted within a broader biophysical framework. We further clarify that possible mechanisms, such as changes in stratification, nutrient resupply, mixing, and seasonal structural reorganization, should be regarded as plausible explanations rather than direct attribution because the present analysis does not explicitly diagnose the underlying physical drivers. This revision avoids interpreting the absence of a statistically significant global mean SCM-depth trend as evidence that the underlying physical controls remained stable.

**6. Discussion**

**Reviewer's comment:**

6.1 L702: Compared with previous studies, the literature review currently presented in the introduction (L126-L141) should be moved here. Please also see my comments 1.6 & 1.7.

**Response:**

We have revised the Introduction and Discussion accordingly. The detailed discussion of previous regional 3D Chl-*a* reconstruction studies has been shortened in the Introduction and revisited in the Discussion when comparing the contribution of the present dataset with previous regional studies. The Discussion now more explicitly emphasizes that previous machine-learning studies have demonstrated the feasibility of reconstructing 3D Chl-*a* structure mainly in regional domains or specific ocean basins, whereas the present study extends Chl-*a* profile reconstruction to the global ocean and further produces a long-term 3D Chl-*a* dataset on regular spatial and temporal grids. This revision highlights the contribution of the present work in terms of global coverage, regular gridded product form, and multiple validation strategies.

**Reviewer's comment:**

6.2 L709: This statement is not placed appropriately, making it reads jump.

**Response:**

We have revised the relevant part of the Discussion to improve logical flow. The statement about larger reconstruction errors near the SCM layer has been integrated into the paragraph discussing methodological contribution and remaining reconstruction challenges. It now serves to explain why layers with strong vertical gradients and deep Chl-*a* maxima remain among the more challenging parts of the water column to reconstruct.

**Reviewer's comment:** 6.3 L720: These features are consistent with previous BGC-Argo based analysis. If feasible, I suggest discussing their similarities and differences with existing GOBM or data-assimilative products.

**Response:**

We agree and have added a qualitative comparison paragraph in the Discussion. The revised text compares the present dataset with satellite ocean-colour products and CMIP-class GOBM outputs. We explain that satellite products provide long-term and high-resolution surface Chl-a records but cannot directly resolve SCM depth, SCM intensity, or vertical Chl-a gradients. CMIP-class GOBMs can simulate Chl-a within physically and biogeochemically consistent frameworks and are expected to represent basin- to global-scale Chl-a structures. However, simulated SCM depth, SCM intensity, and vertical Chl-a gradients may differ among models because of differences in model resolution, ecological process representation, and biogeochemical parameterizations.

The revised Discussion also states that the large-scale SCM pattern reconstructed in this study, with deeper SCMs in tropical and subtropical oceans and shallower SCMs in mid- to high-latitude regions, provides an observation-constrained reference for comparing major spatial organization of subsurface Chl-a structure in model-based products. This addition better addresses the similarities, differences, and limitations of the reconstructed dataset relative to GOBM or model-based products.

**Reviewer's comment:**

6.4 L723: long-term -> multi-time scale? Multiple statements should be rewritten after revising the applications of the data, e.g., L765-L769. Please also see my main concerns.

**Response:**

We have revised the manuscript to moderate the emphasis on long-term trend analysis and to highlight the multi-scale applicability of the dataset. The revised text describes the product as suitable for analyses of climatological seasonal structure, regional trends, interannual variability, and large-scale latitudinal organization. The Conclusions have also been revised. We now state that SCM-depth variability during 2005–2025 is mainly characterized by pronounced regional heterogeneity and a persistent latitudinal background structure, whereas the global annual mean SCM depth does not show a statistically significant long-term deepening or shoaling trend.

**Reviewer's comment:**

6.5 L735: The phrase "observationally sparse regions" is vague—where exactly are these regions? A data distribution figure should be provided, and the specific locations should be more clearly identified here.

**Response:**

We have added a supplementary data-distribution figure (Fig. S1) showing the spatial density of the BGC-Argo Chl-a profiles retained for model training, validation, and testing, together with the Core-Argo inference profiles. In response to the reviewer's concern, we have revised the Discussion to clarify that, in the sentence referring to "observation-sparse regions," this term primarily refers to regions where BGC-Argo Chl-a training profiles are sparse or patchily distributed.

Based on Fig. S1, these regions include parts of the high-latitude and ice-influenced oceans, tropical and subtropical gyre interiors, some marginal seas and coastal-boundary regions, and other areas with limited biogeochemical-profile support. We have also clarified that, during the generation of the gridded product, data-sparse regions refer to areas with less Core-Argo inference-profile support, where the reconstructed field may rely more strongly on larger interpolation radii or nearest-neighbour filling. This distinction has been added to make the uncertainty discussion more specific and to clarify where representativeness errors and interpolation-related uncertainty may be larger.

## Technical comments

**1. Reviewer's comment:**

L17: number should be reported with only two effective digits. Please check the ESSD style requirements to confirm, and if so, update these numbers as well as others throughout the rest of the manuscript.

**Response:**

We have checked the numerical reporting throughout the manuscript. In the main text, we reduced unnecessary numerical precision where appropriate. In tables and validation statistics, sufficient decimal places for reproducibility and consistency across metrics were retained. The final formatted manuscript follows ESSD style requirements.

**2. Reviewer's comment:**

L403-407: The expressions "decidedly more rigorous" and "more demanding" used here are overly subjective; the authors should specify the concrete advantages of their framework.

**Response:**

Thanks, and the subjective wording has been removed. The revised text now objectively describes the validation framework, including random split, year-based split, spatial cross-validation, independent ship-based inferred-profile validation, and final gridded-product validation. We also revised the comparison with previous studies to state that direct one-to-one comparison is limited by differences in study domain, data sources, target variables, and validation design.

**3. Reviewer's comment:**

Figure 3 and description part (L425-L472): I would label subpanels a1, a2, … a5, b1, b2 – b10, which was a little difficult to follow currently.

**Response:**

We have revised Fig. 3 and its caption accordingly. The depth-dependent metrics are now labelled as Fig. 3a1–a5, and the profile comparison and bias panels are labelled as Fig. 3b1–b10. The corresponding text has also been revised to refer to these new subpanel labels.

**4. Reviewer's comment:**

L436-439: Have all correlation coefficients presented here been subjected to statistical significance testing? Are these results statistically reliable?

**Response:**

Statistical significance testing for Spearman's correlation coefficients has been added. The depth-wise Spearman's ρ values are already shown in Fig. 3a as part of the depth-dependent performance evaluation. We revised the corresponding text and figure caption to clarify that Spearman's ρ was calculated for each 20 m depth bin by pooling all valid prediction–observation pairs within the corresponding depth interval, and that the resulting p values were corrected across all evaluated depth bins using the Benjamini–Hochberg false discovery rate (FDR) procedure.

All evaluated depth bins remained statistically significant after FDR correction, with FDR-adjusted $p<0.001$. We also clarified that the uncertainty of $R^2$ was represented using bootstrap-derived 95% confidence intervals. These additions confirm that the correlation patterns shown in Fig. 3a are statistically reliable across the evaluated depth range.

**5. Reviewer's comment:**

L470: Does the reported error range of 0.15–0.35 refer to absolute error? What is the unit of it? If so, it is recommended to clarify the scale and unit of the reported error range, and relative error metrics could be added as supplementary information where appropriate.

**Response:**

We have clarified this point that the reported 0.15–0.35 range refers to profile-level RMSE values calculated for individual validation profiles and reported in $\log_{10}$ Chl-*a* space. We have also prepared supplementary relative-error metrics to support the interpretation of error magnitude.

**6. Reviewer's comment:**

L554: The reported R2 of 0.6252 does not constitute a "reasonably strong explained variance". I recommend revising this characterization to "moderate" or "acceptable given the constraints," as the current phrasing overstates the model's performance.

**Response:**

We agree and have revised the wording of the product-level validation results. The revised text now explicitly notes that the comparison is affected not only by the scale difference between point-based ship observations and monthly 1° gridded fields, but also by the differences between the Argo-based reconstruction system and ship-based CTD observations in observing platform, sampling strategy, spatiotemporal distribution, vertical resolution, and data-processing procedures. These differences introduce additional representativeness and scale-mismatch uncertainties. Therefore, instead of describing the product-level $R^2$ as indicating "reasonably strong explained variance," we now state that the final gridded product shows a high correlation and an acceptable level of agreement under these observational-system and scale constraints.

**7. Reviewer's comment:**

L575: While the text notes a "relatively weak subsurface enhancement" at 80–120 m, this depth range is visually compressed in Figure 5 due to the linear depth scale. As a result, the vertical structure in the upper ocean is difficult to discern. I recommend either adopting a non-uniform depth axis to expand the upper water column or adding an inset panel that zooms in on the 0–200 m layer to better highlight this feature.

**Response:**

The figure has been reorganized into an enlarged 2 × 2 layout with improved labels and colour bars. After comparing different colour scales, we retained the original colour scheme for consistency and interpretability. More importantly, we adopted a non-uniform depth axis for the Chl-*a* depth-related panels to expand the upper water column and make the 10–300 m upper-ocean and subsurface structure, including the weak subsurface enhancement around 80–120 m, easier to identify.

**8. Reviewer's comment:**

L621: "Overall, …" This should belong to a new paragraph.

**Response:**

Thanks. The concluding sentence beginning with "Overall" is now separated more clearly from the preceding discussion of seasonal modulation.

**9. Reviewer's comment:**

L655: The interpretation of the trend results requires greater scientific precision. While the p=0.324 correctly indicates that no statistically significant trend can be detected, this is a statement about the limits of statistical power, not a definitive confirmation that the global mean SCM depth has remained "remarkably stable." Please revise the expression to distinguish between the statistical outcome and the physical reality. For example, replace "global mean SCM depth remains remarkably stable" with: "the global mean SCM depth exhibits no statistically detectable shift," or "the multidecadal evolution doesn't shown a statistically significant global trend", avoiding the overinterpretation that the SCM depth is stable.

**Response:**

We agree and have revised the interpretation of the global mean SCM-depth trend. We removed the phrase "global mean SCM depth remains remarkably stable" and replaced it with the more cautious statement that "*the global mean SCM depth does not show a statistically significant directional shift over the study period*." We also avoided interpreting the statistical result as definitive evidence of physical stability.

**10. Reviewer's comment:**

L677: The phrase "exhibit active multidecadal adjustments" is subjective and lacks quantitative support.

**Response:**

Thanks for the great suggestion and we agree with the reviewer that the description should provide a neutral standpoint. In the revised manuscript, besides the mentioned phases by the reviewer, some subjective text throughout the entire manuscript was examined and replaced with more objective descriptions, such as "*regionally heterogeneous SCM-depth changes*," "*localized deepening or shoaling tendencies*," and "*regional heterogeneity*."

**11. Reviewer's comment:**

L703: The phrase "relatively rigorous validation" is subjective and lacks a clear benchmark. I recommend replacing it with a more objective description of the validation framework. For instance, the authors could refer to the "multi-faceted validation strategy" or "comprehensive validation framework" employed in this study, which includes profile-level, among others.

**Response:**

This wording has been replaced with a more objective description of the validation framework. To avoid another potentially evaluative expression, we used the more neutral phrase "multiple validation strategies." The revised text then specifies the concrete components of the validation framework, including random split, year-based split, spatial cross-validation, independent ship-based inferred-profile validation, and final gridded-product validation.

**Reference:**

Cheng, L., Li, G., Long, S.-M., Li, Y., von Schuckmann, K., Trenberth, K., Mann, M., Abraham, J., Du, Y., Cheng, X., Liu, H., Xu, Z., Liu, M., Peng, Q., Gong, X., Ma, Z., & Yuan, H. (2025). Ocean stratification in a warming climate. Nature Reviews Earth & Environment, 6, 637–655.

Fay, A., & McKinley, G. (2014). Global open-ocean biomes: mean and temporal variability. Earth System Science Data, 6(2), 273–284.