Earth Syst. Sci. Data, 18, 8757–8782, 2026
https://doi.org/10.5194/essd-18-8757-2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Open Access Earth System Science Data

# A global 3D chlorophyll *a* dataset derived from multimodal deep learning reconstruction

**Ziyue Meng[1,2], Yuntao Wang[2], and Lei Han[1]**

[1]Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China
[2]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

**Correspondence:** Yuntao Wang (yuntao.wang@sio.org.cn) and Lei Han (hanlei@ouc.edu.cn)

Received: 22 April 2026 – Discussion started: 5 May 2026
Revised: 9 September 2026 – Accepted: 10 September 2026 – Published: 23 November 2026

**Abstract.** Chlorophyll *a* (Chl *a*) is a key variable for characterizing marine phytoplankton biomass and upper-ocean biogeochemical variability. Existing global products are generally limited to the ocean surface and therefore cannot adequately resolve subsurface vertical structure. In this study, a multimodal Profile-Surface Transformer (PST) framework was developed by integrating Biogeochemical-Argo (BGC-Argo) Chl *a* profiles, Core-Argo temperature-salinity profiles, and satellite-derived surface Chl *a* to reconstruct the three-dimensional (3D) vertical structure of Chl *a* from 10 to 500 m depth. A global monthly mean 3D Chl *a* dataset at 1° spatial resolution for January 2005–June 2025 was subsequently constructed. Evaluation of the dataset demonstrates that PST exhibits robust stability and effective profile reconstruction capability across various generalization scenarios. The coefficients of determination under random split, year-based split, and spatial cross-validation are 0.89, 0.87, and $0.86 \pm 0.01$, respectively. Furthermore, independent external validation using ship-based observations confirms that the model maintains good generalization performance across different observing systems. The constructed dataset successfully reproduces known global patterns of Chl *a* vertical structure and its subsurface chlorophyll maximum (SCM), as well as their seasonal variability. Variability in SCM depth is primarily characterized by regionally heterogeneous redistribution superimposed on a persistent zonal background, rather than by a pronounced monotonic change in the global mean depth. Overall, the constructed long-term 3D Chl *a* dataset integrates surface observations and discrete biogeochemical profiles, delivering a new observationally constrained global product. This dataset is well-suited for studies on the vertical ecological structure of phytoplankton, long-term ocean biogeochemical variability, and the responses of upper-ocean ecosystems to climate change. The dataset is publicly available from Zenodo (Meng et al., 2026a) at https://doi.org/10.5281/zenodo.22726410.

## 1 Introduction

Phytoplankton play a fundamental role in the Earth system, accounting for approximately half of global primary productivity and supporting marine food webs (Mattei and Scardi, 2021). Chlorophyll *a* (Chl *a*) concentration is a key proxy for phytoplankton biomass and primary productivity. Since the late 1990s, continuous observations of surface Chl *a* across the global ocean have been enabled by ocean-color remote sensing (O'Reilly and Werdell, 2019; Sathyendranath et al., 2019; Jackson et al., 2023). These long-term and internally consistent observations have substantially advanced the understanding of seasonal cycles, interannual variability, and regional biogeochemical processes across the global marine ecosystem (Silsbe et al., 2025; Kulk et al., 2020). Among these ecosystem variations, phytoplankton blooms are particularly important because they represent changes in upper-ocean productivity and subsequently influence the marine carbon export (Lomas and Bates, 2004; Siegel et al., 2023).

Satellite-derived Chl *a* products primarily reflect optical signals from the ocean surface and therefore provide only limited insight into the vertical structure of Chl *a* within

the water column. Below the surface, coupled physical-biological processes regulate phytoplankton growth and lead to vertically heterogeneous Chl *a* distributions. In strongly stratified regions, the vertical distribution of Chl *a* within the euphotic zone is jointly controlled by mixed-layer depth (MLD), light attenuation, nutrient supply, and phytoplankton physiological regulation. The subsurface chlorophyll maximum (SCM) is a prominent manifestation of the interplay among these physical and biological processes (Cornec et al., 2021; Xing et al., 2023; Masuda et al., 2021). SCM formation and seasonal evolution are driven by mixed-layer dynamics, light availability, and nutrient resupply (Cullen, 2015; Masuda et al., 2021), whereas it is commonly regulated by phytoplankton photoacclimation under low-light conditions (Behrenfeld et al., 2002). Generally, the SCM develops near the depth where light availability and nutrient supply reach an optimal balance (Yasunaka et al., 2022). Its presence reshapes the vertical structure of Chl *a* within the euphotic zone and influences the vertical distribution of primary production (Marañón et al., 2021; Bouman et al., 2020). Therefore, reliance on surface Chl *a* information alone is insufficient to characterize the vertical structure of marine ecosystems, associated carbon-cycle processes, and their responses to environmental change.

Ship-based observations and long-term time-series stations provide highly accurate vertical biogeochemical information and have played a fundamental role in delineating the oceanic Chl *a* structure and variability. Autonomous profiling observations from the Argo system have expanded the availability of vertically resolved measurements in the global ocean. In particular, Biogeochemical-Argo (BGC-Argo) has greatly improved the sampling of biogeochemical profiles and provides direct observations of vertically resolved Chl *a* together with temperature, salinity and biogeochemical variables (Claustre et al., 2020). However, the spatial coverage and sampling density of BGC-Argo remain limited relative to the scale of global ocean variability. Core-Argo, by contrast, provides much broader hydrographic coverage through dense temperature-salinity profile measurements, but it does not directly observe Chl *a* or other biogeochemical variables (Roemmich et al., 2019). As a result, observations that directly constrain Chl *a* vertical structure remain too sparse to support a systematic assessment of three-dimensional (3D) Chl *a* variability and its regional differences. This limitation is particularly relevant under ongoing environmental change, as changes in stratification, mixed-layer dynamics, and nutrient resupply may modify Chl *a* vertical distribution and SCM structure (Yasunaka et al., 2022). Such changes affect not only surface biomass but also upper-ocean ecological organization and associated biogeochemical cycling (Marañón et al., 2021). Therefore, a globally consistent, long-term 3D Chl *a* dataset is essential for quantitatively assessing changes in the ocean carbon cycle and the feedbacks to the climate system.

To address the spatiotemporal discontinuity of Chl *a* observations, a variety of gap-filling and reconstruction approaches have been developed (Schollaert Uz et al., 2017; Martinez et al., 2020). Traditional methods, including objective analysis, kriging interpolation, and empirical-orthogonal-function-based reconstruction, have been widely used to reconstruct cloud-contaminated satellite ocean-color observations and to generate spatially continuous surface Chl *a* fields (Beckers and Rixen, 2003; Pottier et al., 2006; Li and Heap, 2008; Liu and Wang, 2018). These methods have substantially improved the usability of surface Chl *a* products for large-scale analyses. In recent years, machine learning and deep learning have increasingly emerged as important tools for Chl *a* reconstruction. Early studies largely relied on support vector machines or shallow neural networks to establish statistical relationships between Chl *a* and environmental proxy variables (Richardson et al., 2002; Zhan et al., 2003; Martinez et al., 2020). Subsequent work showed that convolutional neural networks and ensemble learning methods can more effectively capture the nonlinear relationships between Chl *a* and environmental drivers, thereby substantially improving the reconstruction accuracy of surface Chl *a* at regional scales (Park et al., 2019; Roussillon et al., 2023). Because the training data for these approaches are derived mainly from satellite ocean-color observations and related variables, they have generally focused on reconstructing surface Chl *a*. However, as these approaches rely primarily on satellite-derived optical information, they remain limited to surface reconstruction and cannot fully resolve the vertical structure of Chl *a* within the water column. Global Ocean Biogeochemical Models (GOBMs), by contrast, can simulate the 3D distribution of Chl *a* and related ecosystem variables within a dynamically consistent framework and have advanced our understanding of marine biogeochemical variability (Fennel et al., 2022; Séférian et al., 2020). Nevertheless, uncertainties remain in the representation and parameterization of ecological processes, and model outputs can be sensitive to poorly constrained parameter choices and process formulations (Mamnun et al., 2022; Séférian et al., 2020). Although recent model generations have improved in many respects, persistent model-data mismatches and systematic biases remain for several key biogeochemical variables, limiting the observational consistency of simulated marine ecosystem fields (Séférian et al., 2020). Observation-constrained 3D Chl *a* reconstruction therefore requires not only statistical accuracy but also physically plausible vertical structures that remain consistent with upper-ocean environmental conditions.

Owing to the scarcity of vertically resolved observations and the strong regional heterogeneity of Chl *a* vertical structure, developing globally consistent, long-term 3D Chl *a* data products remains challenging. These limitations highlight the need for an observationally constrained global 3D Chl *a* product that can complement both satellite-based surface reconstructions and model-based 3D simulations. Multi-

ple studies have shown that, after quality control and calibration/correction procedures, BGC-Argo Chl *a* data can provide reliable estimates of Chl *a* concentration (Roesler et al., 2017; Schmechtig et al., 2023). These characteristics make BGC-Argo profiles an important observational basis for reconstructing Chl *a* vertical structure. This complementarity between BGC-Argo and Core-Argo provides a feasible approach for transferring Chl *a* vertical structure information from the relatively sparse BGC-Argo network to the much denser Core-Argo hydrographic network.

Recent studies have demonstrated that vertical or 3D Chl *a* structure can be reconstructed using machine-learning frameworks that combine profile observations, satellite information, and hydrographic variables. For example, previous work has estimated SCM features or reconstructed regional 3D Chl *a* structures in the northwestern Pacific and northern Indian Ocean (Chen et al., 2022; Hu et al., 2023; Zhao et al., 2024). Transformer-based models have also been explored for global pigment-profile estimation, including Chl *a* (Yang et al., 2024). Most existing studies remain focused on specific regions, hydrographic regimes, or profile-level estimation; however, a globally consistent, long-term, gridded 3D Chl *a* product remains underdeveloped. The key challenge is therefore not simply to extend regional reconstruction to a larger domain, but to establish a unified framework for integrating sparse biogeochemical profiles, dense hydrographic profiles, and satellite surface information into a standardized global data product.

To address this challenge, this study develops a multimodal Profile-Surface Transformer (PST) framework for reconstructing the vertical distribution of oceanic Chl *a*. During model training, BGC-Argo Chl *a* profiles, collocated Core-Argo temperature-salinity profiles, and satellite-derived surface Chl *a* observations are jointly used to learn the statistical relationships between upper-ocean environmental conditions and Chl *a* vertical structure. The trained model is then applied to the broader Core-Argo profile network to infer Chl *a* vertical distributions at locations where only physical profiles are available. These inferred profiles are subsequently mapped onto a regular monthly grid to construct a global 3D Chl *a* dataset for 2005–2025 over the depth from 10 to 500 m. The resulting product is evaluated using internal validation, spatial and temporal generalization tests, and independent ship-based observations. The resulting dataset is intended to support analyses of Chl *a* vertical ecological structure, seasonal-to-interannual variability, regional long-term changes, and comparisons with model-based biogeochemical products.

## 2 Data sources

### 2.1 BGC-Argo Chl *a* profile observations

The Chl *a* profile observations used for model training and validation in this study were obtained from the global BGC-Argo observing network. The spatial densities of the BGC-Argo profiles retained for model training, validation, and testing are shown together with those of the Core-Argo inference profiles in Fig. B1a and b. BGC-Argo floats are equipped with multiple biogeochemical sensors that provide vertical profile measurements of variables including Chl *a* fluorescence. Because Chl *a* concentrations are often approximately log-normally distributed, a $\log_{10}$ transformation was applied to Chl *a* during model development and for log-domain validation analyses (Glover et al., 2018), and the final gridded product was back-transformed to linear Chl *a* concentrations. A total of 140 818 BGC-Argo Chl *a* profiles collected during 2005–2025 were assembled in this study. For each BGC-Argo profile, the corrected Chl *a* field was extracted from the profile file and used as the target variable for subsequent matching and model training. The extracted profiles were further screened to remove profiles with obvious outliers, insufficient vertical sampling resolution, or unreliable pairing with physical profiles. After quality control, 109 083 high-quality BGC-Argo Chl *a* profiles were retained for model training, validation, and testing. Previous studies have evaluated the quality control and correction of biogeochemical measurements from profiling floats, and have shown that radiometric information can help improve fluorescence-based Chl *a* retrievals (Johnson et al., 2017; Xing et al., 2011). These studies support the use of quality-controlled BGC-Argo Chl *a* profiles as observational constraints for the reconstruction framework. Their annual profile-count variation is further compared with that of the Core-Argo inference profiles in Fig. B1c. These data provide vertical information on Chl *a* distribution from the ocean surface to the interior of the euphotic zone and serve as the core observational constraint for reconstructing the global 3D Chl *a* dataset developed in this study.

### 2.2 Core-Argo temperature and salinity profiles

The temperature and salinity profile data were obtained from the global Core-Argo observing network. In total, 2 975 413 Core-Argo temperature-salinity profiles collected during 2005–2025 were compiled. After preprocessing and quality screening, 2 816 143 valid profiles were retained. Their spatial and temporal distributions are summarized together with those of the retained BGC-Argo profiles in Fig. B1a–c. This dataset is characterized by broad global coverage, a long observation period, and relatively stable vertical sampling, thereby providing an important basis for characterizing the vertical structure of the oceanic physical environment and for subsequent global 3D Chl *a* reconstruction.

### 2.3 MODIS remote-sensing data

Satellite-derived surface Chl *a* concentration products from MODIS Aqua were used as an auxiliary data source in this study. A total of 944 scenes of 8 d-averaged Level-3 standard

products for 2005–2025 were collected and preprocessed following quality control and cloud-coverage screening. To balance the representation of large-scale spatial patterns and regional surface details, the original-resolution data were resampled to a $0.25° \times 0.25°$ grid using area-weighted averaging. These data primarily reflect surface biomass levels and water-column optical characteristics and were used to characterize the biological background of the surface layer and to provide supplementary information for vertical reconstruction. This 0.25° MODIS field was used to extract local surface Chl *a* information for each Argo profile rather than to define the resolution of the final 3D gridded product. Because the vertical reconstruction is ultimately constrained by the spatial density of Argo profiles, monthly sampling support, and interpolation robustness, the final gridded 3D Chl *a* product was generated at a coarser $1° \times 1°$ resolution.

In this study, BGC-Argo Chl *a* profiles were used as the observational constraint on vertical Chl *a* structure, Core-Argo temperature-salinity profiles were used as the primary physical input carrier, and MODIS surface Chl *a* data were used to provide auxiliary information for the upper ocean. Together, these datasets were used for training-sample construction and the subsequent large-scale reconstruction of Chl *a* vertical structure.

## 3 Methodology

### 3.1 Overall reconstruction framework

The overall workflow of the global 3D Chl *a* reconstruction framework includes the training phase, inference phase, and final gridded product generation (Fig. 1). During training, BGC-Argo Chl *a* profiles provide supervisory signals, while matched Core-Argo temperature-salinity profiles and satellite-derived surface Chl *a* observations provide model inputs. During inference, the trained PST model is applied to Core-Argo profiles containing only temperature and salinity observations to infer Chl *a* vertical distributions at broader spatial and temporal coverage.

This framework enables the transfer of Chl *a* vertical-structure information from the sparse BGC-Argo biogeochemical observations to the denser Core-Argo physical profile network, providing the basis for generating the global gridded 3D Chl *a* product.

### 3.2 Data pre-processing

Before model training and inference, observations from different sources were processed through a unified preprocessing workflow to ensure comparability among profiles in terms of time, space, vertical structure, and numerical scale. To construct joint physical-biogeochemical samples for model training and to provide a consistent data basis for subsequent large-scale Chl *a* inference based on Core-Argo profiles, a hierarchical matching strategy was adopted to pair BGC-Argo Chl *a* profiles with Core-Argo temperature-salinity profiles. In the Argo data system, Core-Argo temperature-salinity profiles and BGC-Argo biogeochemical profiles are disseminated separately at the data distribution level (Argo Data Management Team, 2025). Accordingly, floats were first identified on the basis of platform number (WMO ID) to maximize temporal and spatial consistency. This procedure was intended to reduce potential representativeness errors associated with mismatched observations (Janjić et al., 2018). Strict spatiotemporal matching constraints were further imposed so that only sample pairs with an observation time difference of less than 1 d and a geographic separation of no more than 0.05° were retained.

After spatiotemporal matching, the Core-Argo temperature and salinity profiles were interpolated to the pressure levels of the BGC-Argo Chl *a* observations to achieve vertical alignment and construct joint physical-biogeochemical profiles on a common depth coordinate. The successfully matched joint profiles were then subjected to further quality control and vertical reconstruction. First, valid-value screening and outlier removal based on a sliding-window median absolute deviation (MAD) method were applied to the raw profiles. Subsequently, Gaussian process regression (GPR) was used to smooth and interpolate Chl *a* along depth, while temperature and salinity were linearly interpolated onto a common set of standard depth levels, thereby yielding vertically continuous and internally consistent physical-biogeochemical profile structures. For the Core-Argo temperature-salinity profiles used during inference, the same vertical standardization strategy used in the training process was applied. Raw temperature-salinity profile files were first read in batches to extract temperature, salinity, observation time, and geographic location, and profiles with missing metadata, too few valid observations, or obvious outliers were removed. Pressure was then converted to depth according to observation latitude, and temperature and salinity were interpolated onto a common standard depth grid (10–500 m at 10 m intervals). The processed valid profiles were saved in batches to support large-scale model inference and ensure consistency in the vertical dimension of the input samples.

For the MODIS surface Chl *a* data, the original 8 d-averaged Level-3 products were first subjected to quality control and missing-data screening and then resampled to a $0.25° \times 0.25°$ grid using area-weighted averaging. Based on the observation time and location of each Argo profile, the corresponding MODIS surface Chl *a* patch was matched to both the training and inference samples as an auxiliary surface feature input to the model. Through this procedure, the surface biomass background information provided by remote sensing was integrated with Argo profile observations within a unified spatiotemporal framework, thereby strengthening the model's constraint on Chl *a* vertical structure. After the preprocessing steps, both the training and inference samples were standardized in terms of time, space, and vertical co-

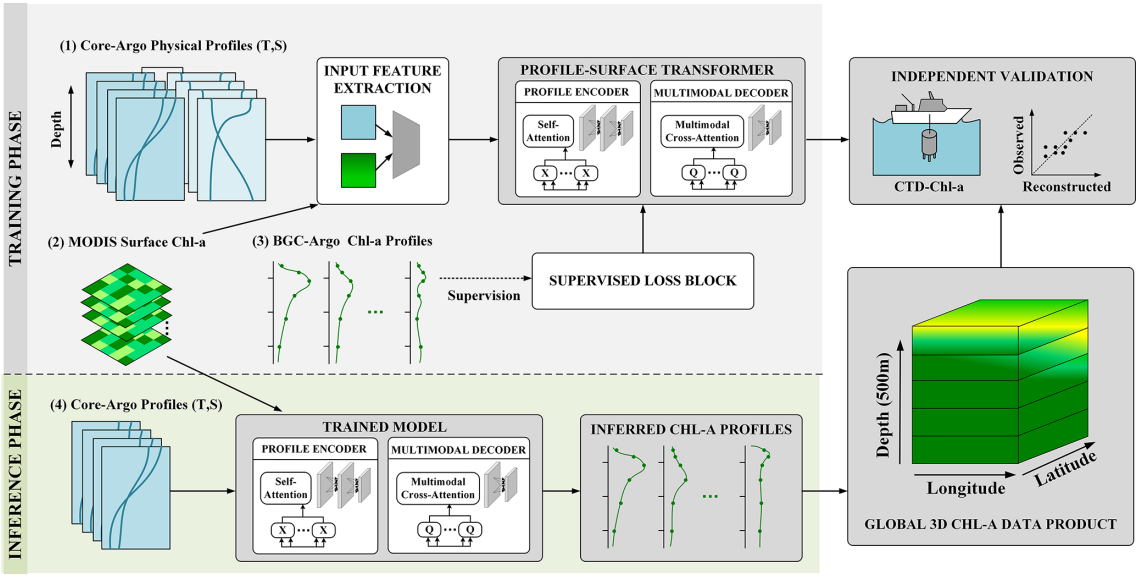



**Figure 1.** Framework of the proposed global 3D Chl *a* reconstruction, including the training phase, inference phase, and final data product generation.

ordinates, thereby providing consistent inputs for subsequent PST model training and global 3D Chl *a* reconstruction.

### 3.3 PST model for reconstruction

A multimodal Transformer model, termed the PST (Fig. 2), was developed to reconstruct vertical Chl *a* profiles on a standard depth grid by jointly using profile information and co-located satellite surface Chl *a* observations. For each sample, the model takes as input a depth-ordered profile sequence together with a matched local MODIS Chl *a* patch extracted at the corresponding time and location.

The profile input includes the original predictors together with an additional approximate density-anomaly feature computed from temperature and salinity. This feature was defined as

$$\sigma_\rho = \rho_{\mathrm{simp}}(T, S) - 1000, \tag{1}$$

where $T$ and $S$ denote temperature and salinity, respectively. The simplified density term was calculated as

$$\rho_{\mathrm{simp}}(T, S) = \rho_{\mathrm{w}}(T) + A(T)\,S + B(T)\,S^{1.5}, \tag{2}$$

where $\rho_{\mathrm{w}}(T)$, $A(T)$, and $B(T)$ are temperature-dependent polynomial terms. The polynomial form follows the temperature-salinity dependence of the one-atmosphere seawater density equation but was simplified here for use as an auxiliary model predictor (Millero and Poisson, 1981; Fofonoff and Millard, 1983). This variable is pressure independent and should not be interpreted as a complete potential-density estimate following EOS-80 or TEOS-10 conventions (Fofonoff and Millard, 1983; McDougall and Barker, 2011). The full polynomial coefficients are provided in Appendix A1. To better represent vertical position, the profile features are further augmented with explicit depth-related terms before entering the main network.

The profile backbone adopts a parallel feature-extraction design. The augmented input is projected into a latent space through a layer-wise linear embedding branch, while a 1D convolution branch extracts local vertical structural information along the depth dimension. A learned gating mechanism derived from the convolutional features is then used to modulate the convolution branch, and the resulting features are fused with the linearly embedded features to form a unified profile representation. Positional encoding is subsequently applied to preserve depth-order information (Vaswani et al., 2017).

To capture vertical structures at different scales, the position-encoded profile representation is passed to two encoders operating at different vertical resolutions. One encoder processes the full depth sequence, whereas the other processes a vertically down-sampled sequence to extract smoother large-scale structural information. The coarse encoded features are interpolated back to the original depth resolution and fused with the full-resolution encoded representation.

For the surface-information branch, the matched MODIS Chl *a* patch is reshaped into a two-dimensional grid and encoded by a 2D convolutional network. The resulting feature map is then converted into fine-scale and coarse-scale surface tokens through multi-scale pooling, followed by normalization and linear projection.

In the decoding stage, learnable depth queries are introduced to represent the target depth levels, in a manner anal-

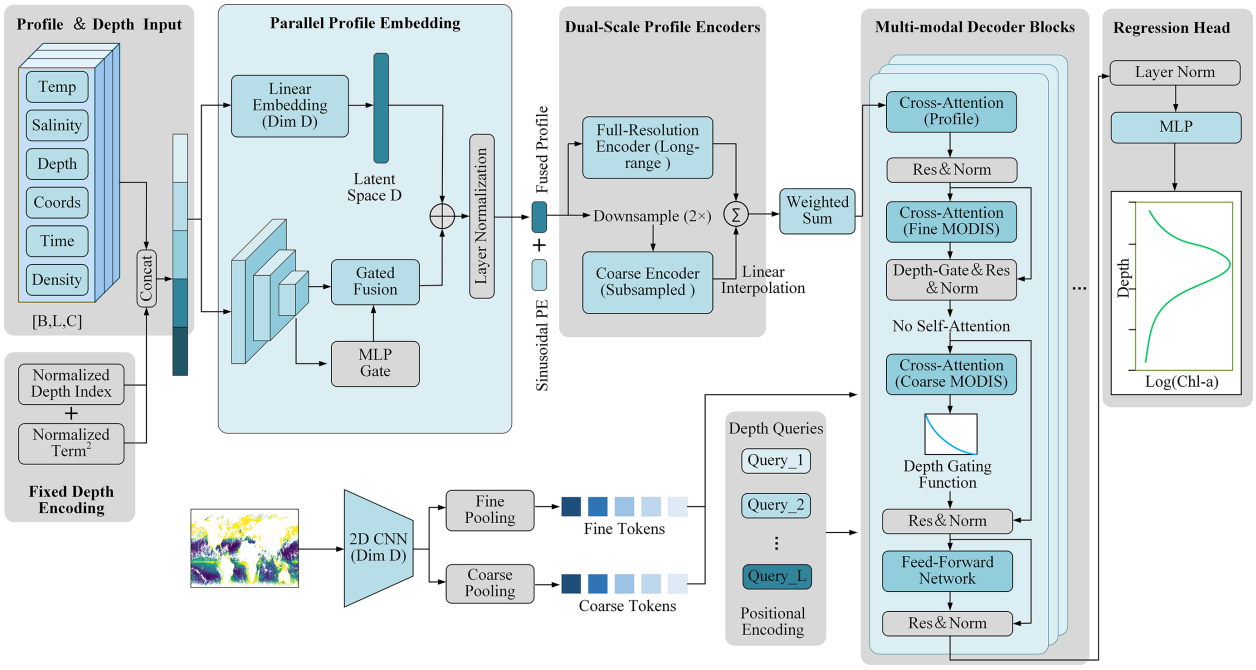



**Figure 2.** Schematic architecture of the PST model, showing profile embedding, profile encoding, multimodal decoding, and Chl *a* profile regression.

ogous to query-based Transformer decoding (Carion et al., 2020). These queries are updated through stacked multimodal decoding blocks. In each block, the queries successively attend to the encoded profile memory, fine-scale surface tokens, and coarse-scale surface tokens, and are then refined by a feed-forward sublayer. Unlike a standard Transformer decoder, no self-attention is applied among the depth queries.

To modulate the direct influence of satellite surface information with depth, the outputs of the fine- and coarse-scale MODIS cross-attention branches were multiplied by a prescribed depth-dependent gate before being added back to the decoder query states. For the $k$th depth level, with $k = 1, \ldots, L$, the normalized depth index was defined as

$$z_k = \frac{k-1}{L-1}, \tag{3}$$

where $L$ is the number of standard depth levels. The gate was then defined as

$$g(z_k) = \exp\left(-\frac{z_k}{\tau}\right), \tau = 0.3. \tag{4}$$

In addition, the fine- and coarse-scale MODIS attention outputs were scaled by branch coefficients $\alpha_{\text{fine}}$ and $\alpha_{\text{coarse}}$, which were initialized as 0.3 and optimized during training. Thus, the exponential depth-decay function was prescribed, whereas the relative strengths of the two MODIS branches were learned from the training data. A sensitivity test was further conducted by comparing the default setting ($\tau = 0.3$) with alternative decay scales and a no-gate configuration. Additional experiments using alternative decay scales and a no-gate configuration showed only marginal differences in reconstruction performance, indicating that the model was not highly sensitive to the prescribed value of $\tau$ (Table A1). The baseline value of $\tau = 0.3$ was therefore retained. This multimodal interaction follows the general cross-attention paradigm of the Transformer architecture (Vaswani et al., 2017), while extending it to multiple memory sources, including encoded profile features and fine- and coarse-scale MODIS surface tokens.

Finally, the decoded depth-wise features are mapped by a regression head to predict $\log_{10}$-transformed Chl *a* at each standard depth level. Before model training, Chl *a* concentrations were transformed to $\log_{10}$ space to reduce the influence of their strong right-skewness and large dynamic range. Accordingly, the model output and loss function were also defined in $\log_{10}$-transformed Chl *a* space. Predicted values were back-transformed to linear Chl *a* concentrations for the generation of the final gridded product. Model training minimizes a composite loss consisting of a depth-weighted masked mean squared error and a curvature-consistency term based on second-order vertical differences. The total loss was defined as

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_{\text{curv}}\mathcal{L}_{\text{curv}}, \lambda_{\text{curv}} = 0.2. \tag{5}$$

Here, the masked MSE term was computed as

$$\mathcal{L}_{\mathrm{MSE}} = \frac{\sum_{b=1}^{B}\sum_{k=1}^{L} m_{b,k} w_k \left(\hat{y}_{b,k} - y_{b,k}\right)^2}{\sum_{b=1}^{B}\sum_{k=1}^{L} m_{b,k} w_k + \epsilon}, \tag{6}$$

where the depth weight was prescribed as

$$w_k = 1 + 2\frac{k-1}{L-1}. \tag{7}$$

Thus, the weight increased linearly from 1 at the shallowest level to 3 at the deepest level. For the curvature-consistency term, the second-order vertical difference was defined as

$$D^2 y_{b,k} = \frac{y_{b,k+1} - 2y_{b,k} + y_{b,k-1}}{\Delta z^2}, \Delta z = 10\,\mathrm{m}. \tag{8}$$

The curvature loss was then calculated as

$$\mathcal{L}_{\mathrm{curv}} = \frac{\sum_{b=1}^{B}\sum_{k=2}^{L-1} m_{b,k-1} m_{b,k} m_{b,k+1} \left(D^2 \hat{y}_{b,k} - D^2 y_{b,k}\right)^2}{\sum_{b=1}^{B}\sum_{k=2}^{L-1} m_{b,k-1} m_{b,k m_{b,k+1}} + \epsilon}. \tag{9}$$

In these equations, $\hat{y}_{b,k}$ and $y_{b,k}$ denote the predicted and observed $\log_{10}$-transformed Chl *a* values for sample $b$ at depth level $k$, respectively; $m_{b,k}$ is the valid-observation mask; $w_k$ is the depth-dependent weight; $B$ is the batch size; $L$ is the number of standard depth levels; and $\epsilon = 10^{-8}$ is used for numerical stability.

The first term improves layer-wise fitting under incomplete observations, while the second term incorporates consistency between the predicted and observed vertical curvature and helps preserve vertical structural continuity. Together, these two terms facilitate both accurate reconstruction and physically coherent vertical profiles.

### 3.4 Training and inference strategy

The PST model was trained using the matched BGC-Core joint profile dataset. Before training, all samples underwent a unified procedure for quality control and feature construction. The model input consisted of depth-ordered temperature-salinity features together with their spatiotemporal encodings, and the output was the $\log_{10}$-transformed Chl *a* concentration at the corresponding depth levels. During training, a valid-observation mask was introduced such that the loss was computed only at depth levels constrained by observations, thereby reducing the influence of missing values on parameter optimization. By minimizing the overall discrepancy between predicted and observed profiles, the model learned the statistical relationship between Chl *a* vertical structure and the physical environment.

To systematically evaluate model performance under overall distributional variation, spatial generalization, and temporal extrapolation, the dataset was divided into training, validation, and testing subsets using three strategies: random split, spatial cross-validation, and year-based split, as summarized in Table 1. In the spatial cross-validation experiment, $K$ denotes the number of spatial folds, with $K = 5$ in this study. Samples were first grouped into fixed $5° \times 5°$ latitude-longitude blocks according to their profile locations, and these blocks were then randomly assigned to five mutually non-overlapping spatial folds using a fixed random seed. In each fold, one complete spatial fold was withheld as the spatially independent test set, while 20 % of the remaining non-test blocks were used for validation and the rest for training. This design was motivated by the fact that random splitting in spatiotemporally correlated datasets may lead to overly optimistic performance estimates, whereas spatially and temporally independent splits are better suited to reflect realistic extrapolation scenarios (Koldasbayeva and Zaytsev, 2025). The specific configuration of the spatial cross-validation also followed recent recommendations regarding spatial blocking strategies in geoscience and ocean remote sensing studies (Stock, 2025).

During model training, the validation set was used for hyperparameter tuning, and an early-stopping strategy was adopted to prevent overfitting. Model performance was evaluated not only using overall statistical metrics but also by examining the distribution of errors across different depth ranges, in order to ensure that the model could reasonably reconstruct both surface and subsurface Chl *a* structures. During inference, the trained model was applied to Core-Argo profiles containing only temperature and salinity observations. This procedure ultimately generated global predictions of Chl *a* vertical distributions with high spatiotemporal coverage, thereby providing the foundational input for the subsequent 3D reconstruction of the dataset.

### 3.5 Construction of the 3D Chl *a* field

The trained model was first applied to all Core-Argo profiles to infer Chl *a* at each standard depth level, yielding profile-by-depth predictions with latitude, longitude, time, and the predicted Chl *a* value. To generate a regular gridded product suitable for climate analysis and visualization, these discrete profile-based predictions were mapped onto a unified latitude-longitude-depth-time grid, from which a global monthly 3D Chl *a* field for 2005–2025 was constructed.

Spatial interpolation was performed using an adaptive-radius Gaussian-weighted scheme based on spherical distance. For each month and depth level, all profile locations and target grid points were first projected onto a unit-sphere coordinate system. For each target grid cell, neighboring profiles were first identified within the maximum candidate radius of 700 km. The smallest radius satisfying the prescribed minimum-neighbor constraint was then selected automatically from the candidate radii of 100, 200, 300, 500, and 700 km, so that local neighborhoods were retained wherever observational support was sufficient. Specifically, if the number of neighboring profiles within a candidate radius reached

**Table 1.** Dataset partitioning strategies.

| Partitioning Strategy | Principle | Training Set | Validation Set | Test Set |
|---|---|---|---|---|
| Random split | Random sampling from the full dataset. | 70 % | 15 % | 15 % |
| Spatial cross-validation | Samples were grouped into fixed $5° \times 5°$ latitude-longitude blocks, which were randomly assigned to $K = 5$ non-overlapping spatial folds. | Samples from the remaining non-test and non-validation spatial blocks. | Samples from 20 % of the non-test blocks. | Samples from one complete spatial fold. |
| Year-based split | Partition strictly by observation year. | 2005–2022 | 2023 | 2024 |

the required threshold (set at no fewer than six in this study), that minimum radius was used. If none of the candidate radii satisfied this condition but at least three neighbors were available within 700 km, a 700 km Gaussian-weighted estimate was generated as a relaxed fallback during the adaptive interpolation stage. The influence of the maximum candidate radius was further evaluated using a random holdout experiment, in which inferred profiles were withheld and reconstructed using the remaining profiles (Table B1). The actual usage frequencies of the candidate radii are summarized in Table B2, and the spatial contribution to the use of the 700 km fallback radius is shown in Fig. B2. These results show that the 700 km radius was not the nominal or dominant interpolation scale, but a limited upper-bound fallback used mainly where smaller radii did not provide sufficient neighboring-profile support. Grid cells with fewer than three neighboring profiles within 700 km were not assigned a Gaussian-weighted estimate at this stage.

This adaptive strategy was designed to balance local representativeness, spatial coverage, and interpolation stability under uneven sampling, consistent with general considerations in environmental spatial interpolation and Argo gridding under nonuniform sampling (Li and Heap, 2008; Li et al., 2017). It follows the general logic of distance-weighted objective analysis and Barnes-type interpolation, in which irregular observations within a radius of influence are mapped onto regular grid points using distance-dependent weights (Barnes, 1964; Udaya Bhaskar et al., 2007; Zürcher, 2023). The Gaussian weighting formulation used here is analogous to Barnes-type interpolation, which estimates a gridded value as a normalized Gaussian-weighted average of irregularly distributed observations (Udaya Bhaskar et al., 2007; Zürcher, 2023). Once the interpolation radius was selected, the interpolated Chl *a* value was computed as a normalized Gaussian-weighted average:

$$\hat{C} = \frac{\sum_i w_i C_i}{\sum_i w_i}, \tag{10}$$

with

$$w_i = \exp\left(-\frac{1}{2}\left(\frac{d_i}{r}\right)^2\right), \tag{11}$$

where $C_i$ is the predicted Chl *a* value of the $i$th neighboring profile, $d_i$ is the spherical distance between the target grid cell and the neighboring profile, and $r$ is the interpolation radius selected for the target grid cell. To reduce the influence of extreme values on the interpolation results, the predicted Chl *a* values were subjected to percentile-based clipping prior to interpolation for each month-depth layer, and a lower bound was imposed on very low values to improve robustness.

An ocean mask derived from vector coastline data was incorporated during interpolation to prevent land areas from being erroneously filled. After interpolation, any remaining missing grid cells within the ocean domain were filled using a nearest-neighbor approach. Therefore, the 700 km radius should be interpreted as the largest candidate radius used in the adaptive Gaussian interpolation stage, whereas the final completion of residual missing ocean grid cells was performed by nearest-neighbor filling. If an entire month-depth layer contained fewer than 50 valid profile points, the adaptive interpolation step was not performed and the layer was retained as missing unless sufficient adaptive-interpolation source values were available for subsequent filling. In addition to the final gridded Chl *a* field, the updated NetCDF product includes ancillary variables describing sampling support, adaptive-radius selection, interpolation failure, ocean-domain coverage, and dependence on nearest-neighbor filling. These variables are described in Appendix B4 and allow users to identify regions, months, and depth levels with different levels of profile support and interpolation dependence. In general, grid cells supported by more neighboring profiles and smaller adaptive radii are more strongly constrained by nearby profile information, whereas cells requiring larger radii, adaptive-interpolation fallback, or nearest-neighbor filling should be interpreted with greater caution. These diagnostics provide practical indicators of relative product reliability but do not constitute a formal decomposition of observational, model, representativeness, and interpolation uncertainties. The resulting 3D Chl *a* dataset is consistent in spatial resolution, temporal continuity, and vertical structure, thereby providing a stable data foundation for marine ecological and biogeochemical studies at both global and regional scales.

### 3.6 Independent ship-based validation data and matching strategy

The independent ship-based CTD observations used for external validation were processed and matched before comparison with the reconstructed results. The ship-based observations were first standardized with respect to time, year, latitude and longitude. Because fluorescence-derived Chl *a* profiles may contain a non-zero deep background signal, a deep-background correction was applied before validation. For each station profile within a cruise, valid raw fluorescence-derived Chl *a* measurements below 300 m were used to estimate the deep-background value. Specifically, when sufficient valid measurements were available below 300 m, the background value was estimated as the median of these deep measurements and was subtracted from the original fluorescence-derived Chl *a* values throughout the corresponding profile. This median-based correction was used to reduce the influence of deep fluorescence background drift and occasional deep spikes. The corrected Chl *a* values were then subjected to the same valid-value screening, cruise-level filtering, and matching procedure used in the external validation.

For inferred-profile validation, the ship-based CTD profiles were matched to the Argo-inferred Chl *a* profiles using constraints on time difference, horizontal distance, and vertical depth. For product-level validation, the processed ship-based observations were matched to the final monthly gridded product by month, grid cell, and depth level over the 10–500 m depth range. For comparison, the ship-based CTD data were vertically averaged within the corresponding depth window to reduce inconsistencies associated with local observational noise and differences in vertical resolution. The use of independent ship-based observations provides an external validation pathway for assessing the reconstructed Chl *a* profiles and gridded product. This validation strategy is consistent with previous studies that used independent bio-optical profiling observations to validate ocean-color products in remote ocean regions (Wojtasiewicz et al., 2018).

## 4 Validation and uncertainty assessment

### 4.1 Internal model validation: Evaluation on the training, validation, and test sets

To assess model stability and robustness under different generalization scenarios, predictive performance was systematically compared using three dataset partitioning strategies: random split, year-based split, and spatial cross-validation (Table 2). The differences in difficulty among these three strategies mainly arise from the degree of independence between the training and test sets and the extent of extrapolation required. In the random split, training and test samples are highly mixed in both time and space, and the evaluation primarily reflects the model's fitting ability under approximately identically distributed conditions. The year-based split requires the model to make predictions for unseen years and therefore primarily tests temporal extrapolation. Spatial cross-validation, by contrast, reserves an entire spatial region for testing, thereby placing greater emphasis on spatial generalization under regionally independent conditions and is thus generally regarded as the most stringent of the three validation strategies.

In both the internal and ship-based validation analyses, $R^2$ denotes the coefficient of determination calculated directly from prediction errors.

Under the random split, the model achieved the highest reconstruction accuracy, yielding an $R^2$ of 0.89, an RMSE of 0.22, and an MAE of 0.158. This confirms that, when the training and test distributions are highly similar, the proposed PST model effectively captures the statistical relationship between vertical Chl *a* structures and the physical environment. These results underscore the strong representational and nonlinear mapping capabilities of the proposed PST model for the reconstruction of complex 3D oceanic structures. Under the more rigorous year-based split, the model maintained high predictive skill on entirely unseen years ($R^2 = 0.87$, RMSE $= 0.24$). The marginal performance drop relative to the random split suggests that the model does not overfit to year-specific anomalies; rather, it learns the underlying long-term statistical regularities of Chl *a* vertical structure, thereby indicating temporal extrapolation capability.

Even under the most challenging of the three validation settings, in which training and test samples were strictly spatially segregated, the model achieved an $R^2$ of $0.86 \pm 0.01$ and RMSE and MAE values of $0.26 \pm 0.01$ and $0.19 \pm 0.01$, respectively. This indicates that performance degradation during spatial extrapolation was limited. Previous studies have shown that ocean remote-sensing and environmental variables exhibit strong spatial autocorrelation; thus, random splits often yield overly optimistic generalization estimates, whereas spatial block cross-validation provides a more realistic assessment of predictive skill in unobserved regions (Stock, 2025). The maintenance of an $R^2$ near 0.86 under this stringent framework substantiates the model's reliability for global-scale extrapolation into observation-sparse oceanic regions.

Crucially, the model's bias remained near zero across all three partitioning strategies, indicating the absence of substantial systematic errors under varying spatiotemporal conditions. This lack of systematic bias is particularly important for the generation of a long-term, internally consistent global 3D Chl *a* product.

Overall, the validation results indicate that the PST framework maintains stable predictive skill across random, temporal, and spatially independent testing conditions. Together with the near-zero bias across validation settings, these results support the use of the PST framework for global 3D Chl *a* reconstruction and large-scale ecosystem analysis.

**Table 2.** Comparison of model test performance under different dataset partitioning strategies.

| Dataset partitioning strategy | RMSE | MAE | Bias | $R^2$ |
|---|---|---|---|---|
| Random split | 0.2249 | 0.1580 | −0.0077 | 0.8923 |
| Year-based split | 0.2433 | 0.1734 | −0.0100 | 0.8739 |
| Spatial cross-validation | $0.2577 \pm 0.0079$ | $0.1863 \pm 0.0069$ | $-0.0089 \pm 0.0043$ | $0.8588 \pm 0.0105$ |

For spatial cross-validation, values are reported as the mean ± sample standard deviation across the five test folds.

The high-resolution depth-stratified statistics at 20 m intervals (Fig. 3a1–a5) show that the model performs consistently across the 10–500 m water column and that all evaluation metrics exhibit clear depth-dependent patterns. RMSE and MAE (Fig. 3a1–a2) reach their maxima at approximately 120–180 m, with values of about 0.29 and 0.21, respectively, and then gradually decrease and level off with increasing depth. This depth range is consistent with the typical distribution depth of the SCM in the global ocean (Mignot et al., 2014; Cornec et al., 2021). The error peak occurs within the SCM layer and its adjacent zones of enhanced vertical gradient, indicating that the model experiences more difficulty when reconstructing regions with the strongest vertical structural variability, whereas errors remain lower in the more vertically homogeneous surface and deep layers. Similar vertical structural characteristics have been systematically documented in multiple BGC-Argo-based studies (Barbieux et al., 2019; Cornec et al., 2021).

Consistent with this pattern, $R^2$ and Spearman's $\rho$ (Fig. 3a4–a5) remain relatively high near the surface, reaching approximately 0.80 and 0.88, respectively. Both metrics decrease gradually with depth within the upper 200 m, followed by a more pronounced drop at around 280–300 m, after which they remain relatively stable in deeper waters. The uncertainty of $R^2$ was represented using bootstrap-derived 95 % confidence intervals, while the statistical significance of Spearman's $\rho$ was evaluated for each 20 m depth bin. Spearman's $\rho$ was calculated for each 20 m depth bin by pooling all valid reconstructed-observed pairs within the corresponding interval. The resulting $p$ values were corrected across all evaluated depth bins using the Benjamini-Hochberg false discovery rate procedure. All correlations remained statistically significant after FDR correction, with adjusted $p < 0.001$. The corresponding Spearman's $\rho$ values ranged from 0.66 to 0.89, and a summary of the significance test is provided in Table C1. This structure is consistent with the physical characteristics of weaker Chl *a* amplitude and a lower signal-to-noise ratio at depth (Arteaga and Rousseaux, 2024). These results indicate that the model performs better in layers with stronger Chl *a* variability, while the correlations in deeper layers with weaker Chl *a* variability remain statistically significant and relatively stable. Bias (Fig. 3a3) remains slightly negative throughout the water column, indicating a weak tendency for the model to underestimate Chl *a* concentration. However, its magnitude remains small at all depths and does not exhibit cumulative growth or marked amplification with increasing depth. This suggests that the bias is more consistent with a weak systematic offset in overall magnitude than with structural instability propagated vertically. In addition, the error bars at all depths are relatively narrow and do not show substantial expansion with depth, further indicating the statistical stability of the model estimates.

At the profile-structure level, the mean predicted profile agrees closely with the mean observed profile across all depth ranges (Fig. 3b1–b5). Within 10–200 m, the predictions accurately reproduce the overall pattern of decreasing Chl *a* with depth together with the mid-depth enhancement, which is a typical vertical structure widely observed in the global ocean (Mignot et al., 2014; Cornec et al., 2021). The shaded envelopes in the upper row (Fig. 3b1–b5) represent the mean ± 1 standard deviation and reflect the range of variability in the observations and predictions. The predicted standard deviation envelope is similar in width to that of the observations, particularly within the 100–200 m interval, where the model reasonably captures the increase in variability associated with enhanced vertical gradients. This indicates that the model not only reproduces the mean structure but also captures the dispersion of the profiles with good fidelity.

The lower row (Fig. 3b6–b10) shows the mean bias profile together with its 95 % confidence interval. Across all depth ranges, the biases remain distributed around zero, and the confidence interval does not increase systematically with depth. No anomalous amplification is observed in the mid-depth layers, suggesting that the model does not introduce obvious structural instability in either the SCM region or the deeper ocean. At the global scale (Fig. 3c), the spatial distribution of profile-level RMSE is relatively uniform overall, without any large-scale zonal or meridional systematic bias. Areas of comparatively higher error are scattered mainly in the high-latitude Southern Ocean and in several dynamically complex regions, such as subtropical frontal zones and areas adjacent to western boundary currents. These regions are characterized by intense frontal activity, water-mass contrasts, and enhanced surface gradients, which may increase spatial heterogeneity and reconstruction difficulty (Janjić et al., 2018; Yang et al., 2023).

Even in these regions, most profile-wise RMSE values, calculated separately for individual validation profiles, remained within 0.15–0.35 in $\log_{10}$-transformed Chl *a* space.

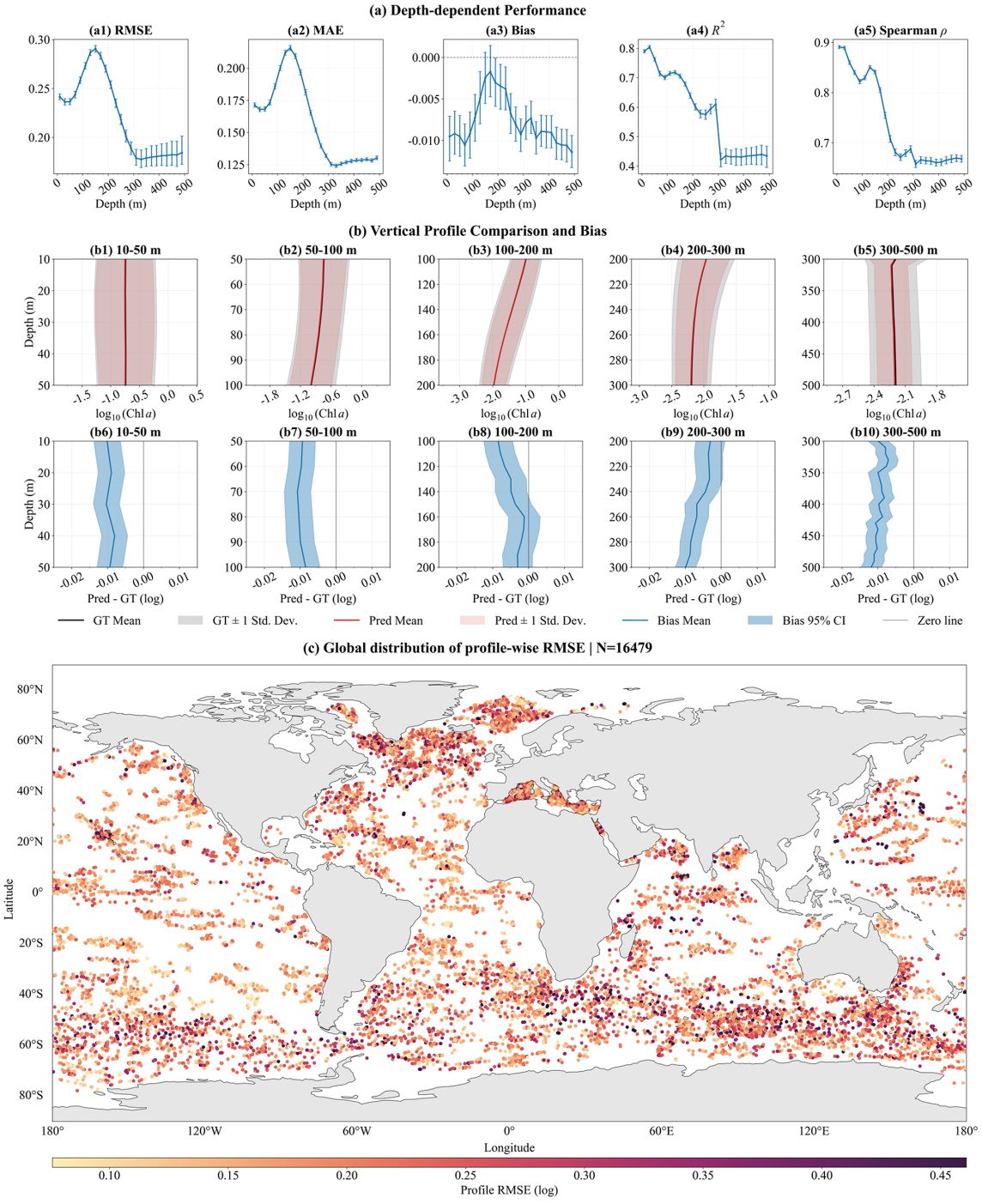



**Figure 3.** Reconstruction performance across depth, vertical profile comparisons in different depth intervals, and the global distribution of profile-level reconstruction errors. **(a1–a5)** Depth-dependent metrics of the inferred Chl *a* profiles, including RMSE, MAE, bias, $R^2$, and Spearman $\rho$. Error bars in panels **(a1)**–**(a5)** indicate 95 % bootstrap confidence intervals. **(b1–b10)** Comparisons of reconstructed and observed vertical structures within five depth ranges (10–50, 50–100, 100–200, 200–300, and 300–500 m). The upper row shows the mean profiles of observations and reconstructions together with their ±1 standard deviation envelopes, and the lower row shows the corresponding depth-dependent bias (Pred-GT) with its 95 % confidence interval. **(c)** Global distribution of profile-wise RMSE values for all evaluated profiles, illustrating the spatial variability of reconstruction errors.

These values represent log-domain error magnitudes and are not expressed in physical concentration units of $mg\,m^{-3}$. Supplementary profile-wise relative errors calculated in the original linear Chl *a* concentration domain are provided in Appendix C2 and Table C2.

Taken together, the depth-stratified statistics (Fig. 3a1–a5), the consistency of the mean profile structure (Fig. 3b1–b5), the mean bias profiles with 95 % confidence intervals (Fig. 3b6–b10), and the global spatial distribution of reconstruction errors (Fig. 3c) demonstrate that the proposed PST model exhibits stable generalization performance across different depth layers and oceanic regions. The model maintains reasonable error levels in key vertical-structure zones such as the SCM, while also preserving stable performance in deeper weak-signal layers, thereby supporting the internal validation of the construction of a long-term, globally consistent 3D Chl *a* dataset.

### 4.2 External validation: Independent generalization test based on ship-based observations

To further evaluate the model's generalization ability under a completely independent observing system, ship-based CTD observations were introduced to externally validate the Argo-inferred Chl *a* profiles. Figure 4a shows the spatial distribution of the cruises used in the external validation, together with the annual distribution of retained validation cruises. After cruise-level quality filtering, 135 cruises were retained, yielding 294 471 matched prediction-observation pairs. These cruises are distributed across the majority of oceanic regions, including the Pacific, Atlantic, Indian, and Southern Oceans, thereby providing good spatial representativeness. The retained validation samples also provide temporal coverage for most of the reconstruction period, with matched samples available for 2005 and 2007–2024. No retained matched validation samples are available for 2006 or 2025 after the cruise-level filtering and matching procedure. Therefore, the external validation provides broad spatial coverage and near-continuous temporal support across the study period.

The ship-based observations are independent of the model training and internal testing process in terms of space, time, and observing platform and therefore provide an external assessment of model transferability under real ocean-observing conditions. As an important component of the global ocean observing system, the complementary relationship between the BGC-Argo network and traditional ship-based observations has been systematically documented in previous studies (Claustre et al., 2020). The preprocessing and matching strategy for the ship-based observations is described in Sect. 3.6.

The validation statistics were evaluated in both linear and $\log_{10}$-transformed Chl *a* spaces because these two domains emphasize different aspects of model performance. Absolute-error metrics, including RMSE, MAE, and bias, were reported in linear Chl *a* space because they directly quantify the error magnitude in physical concentration units. Pearson's correlation coefficient ($R$) and the coefficient of determination ($R^2$) were reported in $\log_{10}$-transformed Chl *a* space because Chl *a* spans several orders of magnitude and the PST model was trained in $\log_{10}$-transformed space. Under this evaluation strategy, the inferred profiles show good agreement with the ship-based observations. In linear Chl *a* space, the RMSE, MAE, and bias are 0.16, 0.06, and $-0.02\,mg\,m^{-3}$, respectively. In $\log_{10}$ Chl *a* space, the corresponding $R$ and $R^2$ are 0.89 and 0.78, respectively. These results indicate that the inferred profiles have relatively small absolute errors in physical concentration units and show strong association and good predictive agreement between the $\log_{10}$-transformed predictions and ship-based observations across a wide concentration range.

The binned calibration curve lies close to the overall 1 : 1 line (Fig. 4b), indicating that the model preserves good correspondence between predictions and observations across different concentration ranges. The residual distribution is approximately unimodal and centered near zero (Fig. 4c), suggesting that the model has only a small overall bias and that the extrapolated results are relatively stable. Given that ship-based observations are often concentrated in coastal zones, frontal regions, and high-biomass waters and tend to exhibit larger spatial heterogeneity, this level of agreement indicates that the model can preserve physically meaningful statistical relationships between the input environmental variables and Chl *a* vertical structure across different observing systems.

It should be noted that the objective of external validation is not to achieve point-by-point agreement with ship-based observations but rather to assess whether the model exhibits systematic bias under an independent observing system. Similar validation strategies based on independent datasets have been used in machine-learning studies of marine bio-optical parameter retrieval and have been shown to be effective for evaluating model generalization ability (Sauzède et al., 2016). Because the ship-based CTD observations and the Argo-based inferred profiles belong to different data systems, objective differences inevitably exist with respect to observing platform, sampling strategy, spatiotemporal distribution, vertical resolution, and data-processing procedures. These differences introduce additional representativeness error and matching uncertainty. Under such conditions, the external validation results indicate that the inferred profiles are generally consistent with independent ship-based observations, which is particularly important for long-term 3D Chl *a* reconstruction. The independent generalization test based on ship-based CTD data demonstrates that the proposed method can reasonably reconstruct the background structure of Chl *a* vertical distributions under a fully independent observing system. In a statistical sense, the model maintains good stability and consistency, further supporting its applicability to global 3D Chl *a* reconstruction.

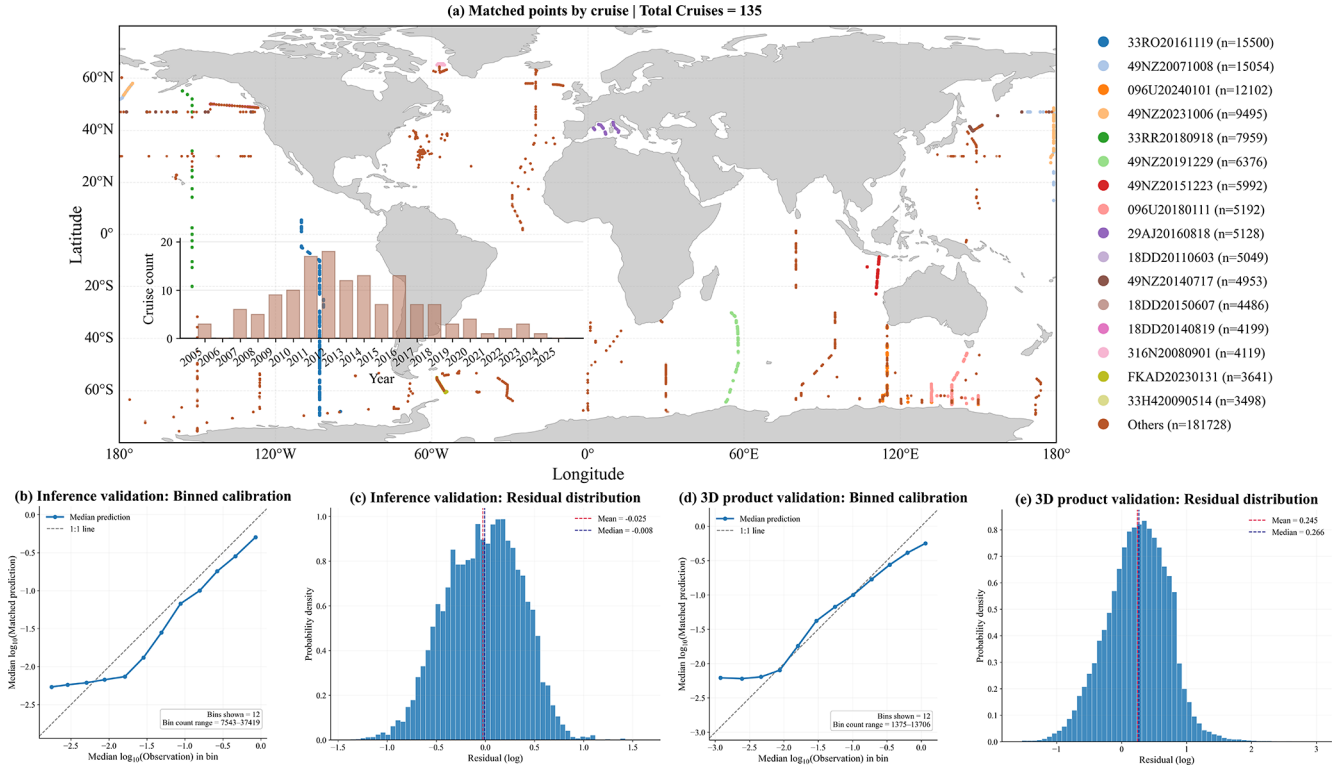



**Figure 4.** Independent external validation of the inferred profiles and the 3D interpolated product using ship-based CTD observations. **(a)** Global distribution of matched validation points by cruise. The inset shows the annual distribution of retained validation cruises during 2005–2024. **(b, c)** Binned calibration and residual distribution for inferred-profile validation. **(d, e)** Binned calibration and residual distribution for 3D product validation. In panels **(b)** and **(d)**, error bars show 95 % bootstrap confidence intervals for the median prediction in each bin. In panels **(c)** and **(e)**, residuals are the differences between $\log_{10}$-transformed predictions and observations.

### 4.3 External validation and reliability diagnostics of the interpolated product

To further assess the reliability of the constructed 3D Chl *a* product on a regular spatiotemporal grid, the independent ship-based CTD observations from the cruises shown in Fig. 4a were also used to externally validate the interpolated results. Unlike the validation of inferred profiles in Sect. 4.2, the comparison here targets the final monthly gridded product. Accordingly, the emphasis is on evaluating the ability of the product to reproduce the background distribution of Chl *a* and its spatial variability. In addition to this external validation, interpolation and sampling-support diagnostics described in Sect. 3.5 are provided as ancillary variables in the NetCDF product to support user assessment of product reliability. The preprocessing of the ship-based data was generally the same as that described in Sect. 3.6. Validation was performed over the 10–500 m depth range, and comparisons were made after matching by month, grid cell, and depth level. To reduce the influence of extreme high-latitude sampling uncertainty, product-level validation was restricted to regions equatorward of 60° latitude. This procedure yielded 70 720 matched product–observation comparison units.

The binned calibration relationship remains generally close to the 1 : 1 line (Fig. 4d), indicating that the product preserves a pattern of variation consistent with the ship-based observations across different concentration ranges. The residual distribution (Fig. 4e) is unimodal, suggesting that the product-level errors are relatively concentrated overall and do not exhibit obvious anomalous offsets.

As in the profile-level validation, the product-level statistics were evaluated by emphasizing different metrics in the domains where they are most interpretable. RMSE, MAE, and bias were reported in linear Chl *a* space to quantify the absolute error of the gridded product in physical concentration units, whereas Pearson's $R$ and $R^2$ were reported in $\log_{10}$-transformed Chl *a* space to evaluate association and predictive agreement between the transformed predictions and observations across the broad dynamic range of Chl *a*. Under this evaluation strategy, the 3D product yielded an RMSE of 0.23 mg m$^{-3}$, an MAE of 0.08 mg m$^{-3}$, and a bias of $-0.01$ mg m$^{-3}$ in linear Chl *a* space. These values indicate that the final gridded product has a relatively small average absolute error and no substantial systematic offset in physical concentration magnitude. In $\log_{10}$-transformed Chl *a* space, the product yielded $R = 0.84$ and $R^2 = 0.63$, indicating a

strong association and meaningful predictive agreement between the transformed predictions and ship-based observations despite the additional smoothing and representativeness errors introduced by monthly gridding and spatial interpolation.

Compared with the validation of inferred profiles in the previous subsection, the interpolated 3D product has undergone further aggregation and resampling onto a monthly, 1° regular grid with standardized depth levels. Although this procedure improves spatiotemporal continuity and practical usability, it inevitably introduces additional smoothing, scale-mismatch, and representativeness errors. Furthermore, because the regular gridded product and the ship-based CTD observations belong to different data systems, differences in sampling strategy, temporal scale, spatial representativeness, and vertical resolution must be considered when interpreting the validation results. Under these conditions, the 3D product constructed in this study shows reasonably low absolute errors and remains broadly consistent with independent ship-based observations in representing the background distribution and spatial variability of Chl *a*. This result suggests that, despite the additional uncertainties introduced by spatial interpolation and monthly gridding, the proposed 3D Chl *a* product remains suitable for large-scale analyses.

## 5 Examples of spatiotemporal variation patterns based on the reconstructed data

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. The SCM depth was diagnosed operationally as the depth of the maximum Chl *a* concentration within the 10–200 m water column. This definition follows a maximum-depth diagnostic commonly used in large-scale Chl *a* maximum analyses (Yasunaka et al., 2022), though it does not apply an additional DCM-occurrence classification. Therefore, shallow values in high-latitude regions during winter or strongly mixed regimes should be interpreted as near-surface or upper-ocean Chl *a* maximum depths rather than necessarily as distinct SCMs.

### 5.1 Global horizontal distribution of climatological SCM depth

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).

### 5.2 Global mean month-depth seasonal characteristics

The global month-depth mean structure shows that the reconstructed 3D Chl *a* dataset captures the main seasonal characteristics of the global mean vertical distribution of Chl *a* (Fig. 5b). Overall, Chl *a* is mainly concentrated within the upper euphotic layer, and a relatively weak subsurface enhancement can be identified at approximately 80–120 m. In deeper waters below 200 m, Chl *a* concentrations remain near zero throughout the year and exhibit little seasonal variability. This indicates that seasonal changes in Chl *a* at the global scale are primarily confined to the euphotic zone and the shallow layers immediately beneath it, while subsurface structure remains an important component of the global vertical Chl *a* distribution. Overall, this result shows that the reconstructed dataset can describe the global mean seasonal cycle and vertical distribution of Chl *a* within a continuous monthly 3D framework.

### 5.3 Climatological latitude-depth structure of reconstructed Chl *a*

The climatological latitude-depth section further reveals the zonal differentiation of Chl *a* vertical structure in the global ocean (Fig. 5c). In tropical and subtropical regions, the SCM is generally located at deeper depths, with peak depths mostly ranging from 80 to 120 m. In contrast, in mid- and high-latitude regions, SCM depth becomes markedly shallower, typically occurring at 40–80 m and in some areas even closer to the surface. This pattern reflects long-term mean differences among latitude bands in light conditions, water-column stratification, and nutrient supply. Previous studies have shown that oligotrophic low-latitude oceans commonly exhibit deeper SCMs, whereas high-latitude waters are more likely to display shallow subsurface maxima or even surface maxima, which is consistent with the latitudinal structure captured by this dataset (Cornec et al., 2021; Yasunaka et al., 2022).

Mechanistically, this latitudinal contrast can be interpreted in terms of photoacclimation, stratification, nutricline depth, and light availability (Yasunaka et al., 2022; Masuda et al., 2021). Global observational studies have further shown that the zonal gradient from deeper Chl *a* maxima at low latitudes to shallower Chl *a* maxima at high latitudes is one of the most robust features of Chl *a* vertical structure (Yasunaka et al., 2022). These results show that the dataset can identify the structural transition from deeper low-latitude Chl *a* maxima to shallow high-latitude Chl *a* maxima across the global ocean, thereby supporting analyses of latitudinal differences in phytoplankton vertical ecological niches.

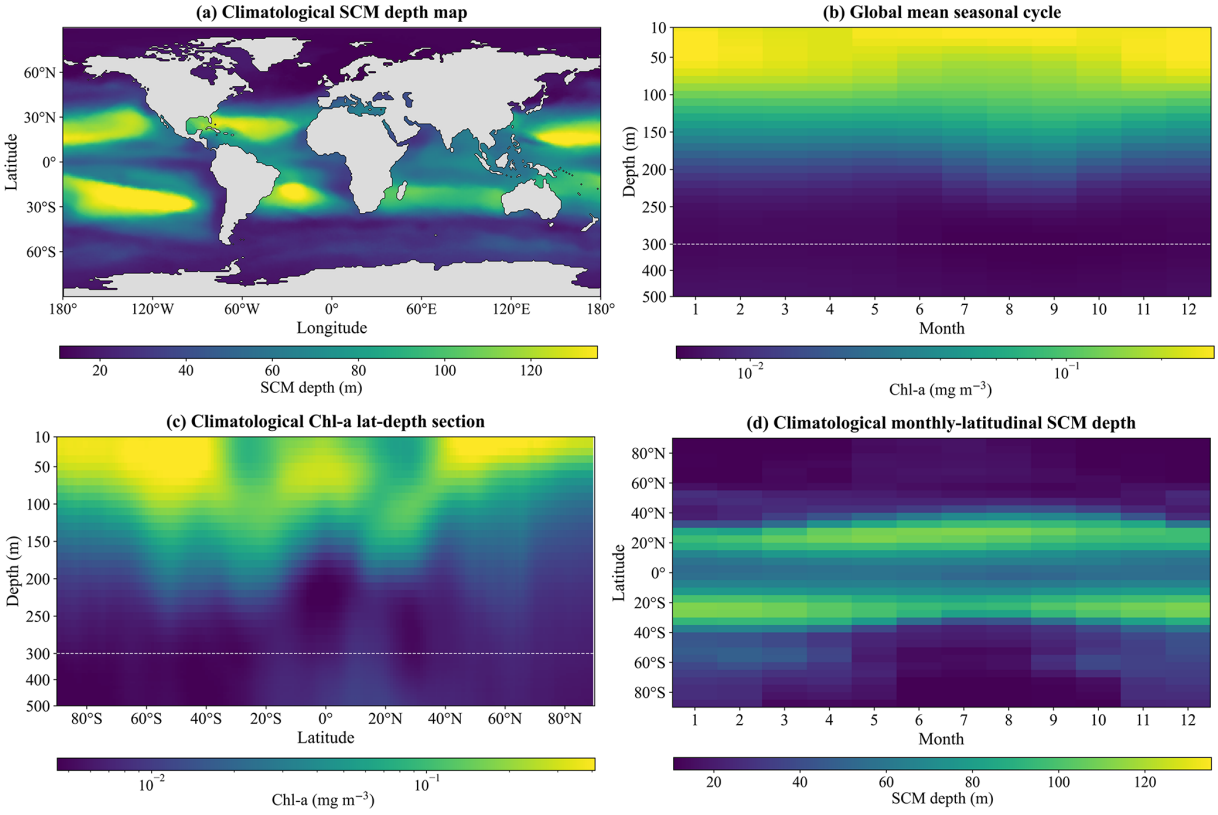



**Figure 5.** Climatological characteristics of reconstructed Chl *a* and SCM depth. **(a)** Global horizontal distribution of climatological SCM depth. **(b)** Global seasonal cycle of reconstructed Chl *a* as a function of month and depth. **(c)** Climatological latitude-depth section of reconstructed Chl *a*. **(d)** Climatological monthly-latitudinal SCM depth. In panels **(b)** and **(c)**, the dashed line marks a depth of 300 m. Below this depth, the vertical plotting scale is compressed.

## 5.4 Monthly-latitudinal seasonal evolution of global SCM depth

The month-latitude structure further reveals the seasonal distribution characteristics of SCM depth in the global ocean (Fig. 5d). Overall, SCM depth exhibits a clear zonal pattern: it is generally deeper in subtropical regions, while equatorial regions show intermediate to relatively shallower SCM depths and mid- to high-latitude regions are characterized by shallower SCM depths. This indicates a stable latitudinal gradient in global SCM depth. This pattern is broadly consistent with global analyses of Chl *a* maximum depth, which show deeper maxima in tropical and subtropical oceans and shallower maxima in mid- and high-latitude waters (Yasunaka et al., 2022).

At the same time, SCM depth still varies with month within individual latitude bands, and seasonal fluctuations are relatively more pronounced in some subtropical and mid- to high-latitude regions. This indicates that the spatial distribution of SCM depth is not static but is seasonally modulated on top of a stable zonal background structure. This seasonal modulation is consistent with previous global observational studies showing that SCM depth varies with latitude and season while maintaining a recognizable large-scale zonal organization (Yasunaka et al., 2022). Overall, Fig. 5d shows that the reconstructed dataset reproduces both the large-scale latitudinal gradient of SCM depth and its seasonal modulation.

## 5.5 Regional trends and interannual variability in SCM depth during 2005–2025

After examining the mean seasonal and spatial structures of reconstructed Chl *a* and SCM depth, the product was further used to illustrate regional trends and interannual variability in SCM depth during 2005–2025 (Fig. 6). Using the resulting SCM-depth fields, global gridded linear trends, the spatially averaged global time series, and zonally averaged latitude-year distributions were derived (Fig. 6a–c). Because observational data are available only through June 2025, the annual mean statistics in Fig. 6b and c were computed using complete calendar years only (2005–2024) to avoid seasonal bias, whereas the trend estimation in Fig. 6a incorporated the full extent of the available 2005–2025 series.

Spatially, SCM depth trends exhibit pronounced regional heterogeneity rather than a globally uniform shift (Fig. 6a). Relatively distinct deepening tendencies are observed along

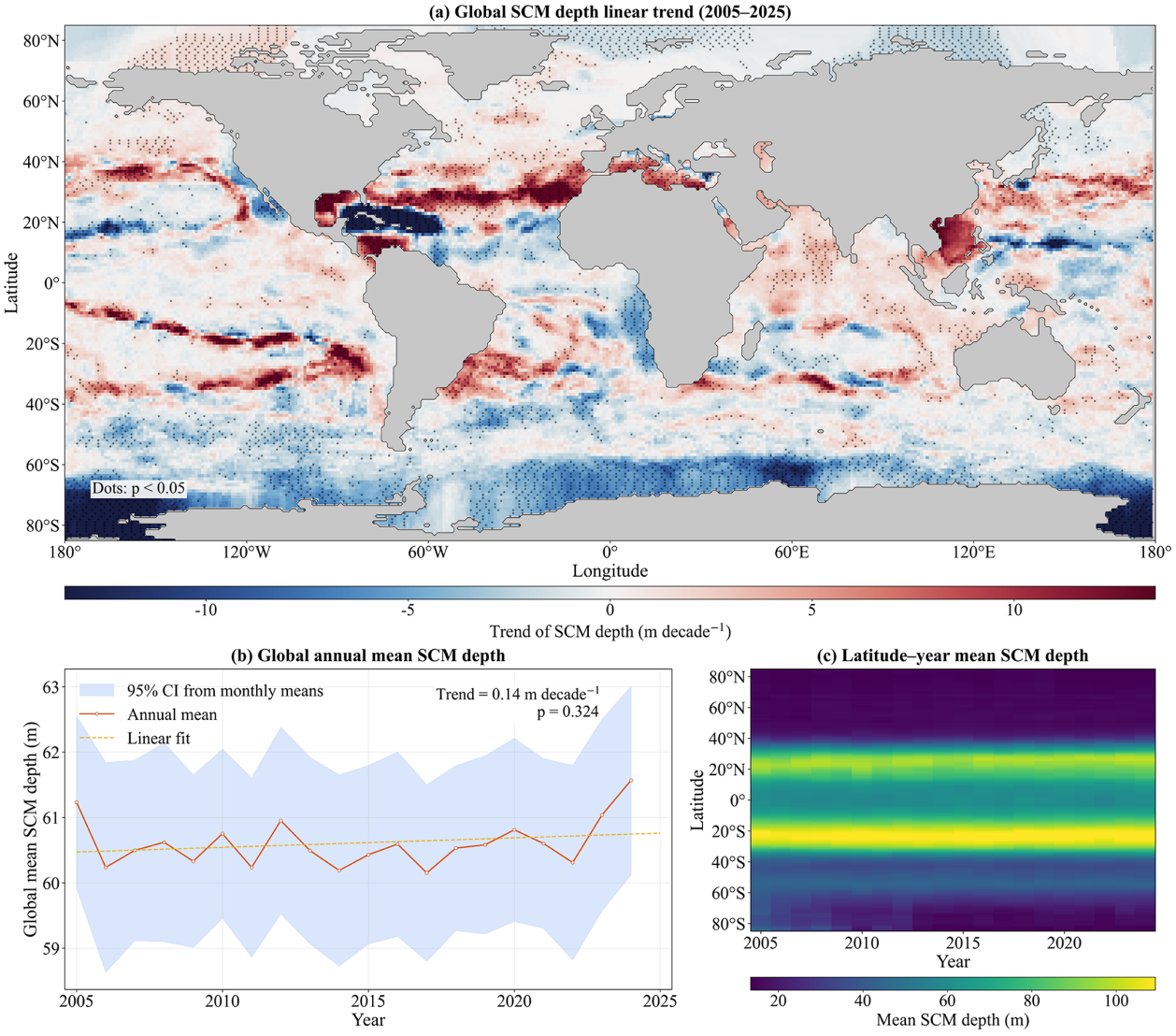



**Figure 6.** Regional trends and large-scale structure of SCM depth (2005–2025). **(a)** Global gridded linear trends of monthly SCM depth (m per decade). Stippling denotes statistically significant trends ($p < 0.05$). **(b)** Time series of the global annual mean SCM depth and its linear fit. Shading represents the 95 % confidence intervals estimated from the monthly means. **(c)** Latitude-year distribution of the zonally averaged annual mean SCM depth. Note: Panel **(a)** uses the full available time series (2005–2025), whereas panels **(b)** and **(c)** are restricted to complete calendar years (2005–2024) to avoid seasonal bias.

the peripheries of subtropical gyres, within certain frontal zones, and across transition regions adjacent to eastern boundary currents. Conversely, patchy shoaling or weak deepening signals are observed near the equator, in some upwelling-influenced regions, and across parts of the high-latitude Southern Ocean. Previous BGC-Argo syntheses have shown that the seasonal dynamics of SCM depth are highly regime-dependent, with systematic regional differences in typical depths and driving mechanisms (Cornec et al., 2021). Therefore, the spatial patterns observed here should not be interpreted as a globally coherent directional drift, but rather as regionally heterogeneous SCM-depth changes across diverse ecological and dynamical environments.

Globally, the annual mean SCM depth fluctuates within a narrow band of approximately 60.0–61.5 m during 2005–2024 (Fig. 6b). The fitted long-term trend is weak and statistically non-significant (0.14 m per decade, $p = 0.324$; Fig. 6b), indicating that interannual variability is more prominent than a monotonic global-mean trend over the analyzed period. Thus, although localized regions exhibit deepening or shoaling tendencies, the global mean SCM depth does not show a statistically significant directional shift over the study period.

The latitude-year distribution mainly highlights the persistence of the large-scale zonal organization of SCM depth during 2005–2024, rather than pronounced year-to-year changes

(Fig. 6c). Subtropical regions remain the regions with the deepest SCM depths, especially in the southern subtropics around 20–30° S, whereas equatorial regions exhibit intermediate depths and mid- to high-latitude oceans are generally characterized by shallower upper-ocean Chl *a* maximum depths, with polar and subpolar regions often showing near-surface maxima shallower than 40 m. This configuration is consistent with previous global observational benchmarks (Yasunaka et al., 2022). This persistent zonal pattern indicates that the reconstructed dataset captures the stable large-scale latitudinal organization of SCM depth.

Taken together, the results shown in Fig. 6a–c indicate that the diagnosed upper-ocean Chl *a* maximum depth during 2005–2025 is characterized by pronounced regional heterogeneity and interannual fluctuations superimposed on a persistent zonal background structure. The observed regional heterogeneity and persistent latitudinal structure are broadly consistent with previous global and BGC-Argo-based studies (Cornec et al., 2021; Yasunaka et al., 2022). Overall, Fig. 6 indicates regionally heterogeneous SCM-depth trends, a weak and statistically non-significant global-mean trend, and a persistent latitude-dependent background structure. These results show that the reconstructed 3D Chl *a* dataset can support multi-scale analyses of SCM-depth variability, while the weak and statistically non-significant global-mean trend should be interpreted together with regional heterogeneity rather than as evidence of a spatially uniform ecological response.

## 6 Discussion

This study develops a global 3D Chl *a* dataset covering 2005–2025 and shows that, under the constraint of BGC-Argo Chl *a* profiles, a data-driven framework combining Core-Argo temperature-salinity profiles and satellite-derived surface Chl *a* information can be used to reconstruct vertically resolved Chl *a* structure at the global scale. The validation indicates that the PST framework maintains good predictive skill under random, year-based, and spatially independent partitioning scenarios, and also shows reasonable agreement with independent ship-based observations. These results suggest that the framework provides not only good predictive accuracy under conventional sample partitioning, but also useful transferability under more stringent spatiotemporal extrapolation conditions.

Compared with previous studies, the main contribution of the present work lies in the combination of global scope, regular gridded product form, and multiple validation strategies. Former machine-learning studies have demonstrated the feasibility of reconstructing 3D Chl *a* structure, but most have focused on regional domains or specific ocean basins (Chen et al., 2022; Hu et al., 2023; Zhao et al., 2024). By contrast, the present study extends Chl *a* profile reconstruction to the global ocean and further produces a long-term 3D Chl *a* dataset with regular spatial and temporal grids. At the same time, the concentration of model errors near 120–180 m indicates that the layers associated with strong vertical gradients and deep Chl *a* maxima remain among the most challenging parts of the water column to reconstruct, even though the model retains useful explanatory power in this depth range.

At the data-product level, the reconstructed global month-depth structure, latitudinal differences, and SCM-depth pattern are broadly consistent with previous global observational studies. Seasonal Chl *a* variability is concentrated mainly within the euphotic zone and nearby subsurface layers, whereas the diagnosed upper-ocean Chl *a* maximum depth exhibits a clear latitudinal structure, with deeper values in tropical and subtropical oceans and shallower values in mid- to high-latitude regions. These spatial features are consistent with previous BGC-Argo-based and global observational analyses (Cornec et al., 2021; Yasunaka et al., 2022). Their interpretation is also consistent with the current mechanistic understanding that SCM-depth variability and subsurface Chl *a* structure are influenced by light attenuation, nutrient supply, stratification, and phytoplankton physiological regulation, including photoacclimation (Cornec et al., 2021; Yasunaka et al., 2022; Masuda et al., 2021). Beyond reproducing these previously recognized large-scale structures, the present dataset extends them into a continuous long-term gridded product that can be tracked and compared across time, depth, and space.

Compared with satellite ocean-color products and CMIP-class GOBM outputs, the present dataset provides a complementary, observation-constrained 3D reference for subsurface Chl *a* structure. Satellite ocean-color records provide long-term, high-resolution global 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 they are expected to represent basin- to global-scale features of Chl *a* structure. However, the representation of subsurface Chl *a* structure, including 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 (Fennel et al., 2022; Mamnun et al., 2022; Séférian et al., 2020). In this context, the large-scale SCM-depth pattern reconstructed in this study, 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. By integrating BGC-Argo Chl *a* profiles, Core-Argo hydrographic information, and satellite surface Chl *a* observations, the present dataset provides a global monthly 3D Chl *a* field constrained by both vertical profiles and surface ocean-color information. Therefore, this dataset can be used to compare key subsurface ecological features, including SCM depth, SCM intensity, vertical Chl *a* gradi-

ents, latitudinal differences, and the seasonal persistence of subsurface Chl *a* structure, thereby helping to fill the gap between surface-focused satellite records and process-based model simulations. The regular global coverage of the product also provides a basis for future biome-, superbiome-, and basin-scale analyses of Chl *a* vertical structure and its spatiotemporal variability (Fay and McKinley, 2014).

A further contribution of this study lies in using the reconstructed product to examine SCM-depth variability during 2005–2025. The key observable feature is not a statistically significant monotonic shift in the global mean SCM depth, but the coexistence of regionally heterogeneous trends and a persistent latitude-dependent background structure. This indicates that localized SCM-depth changes can be obscured by global averaging and highlights the value of a continuous 3D product for examining regional trends and large-scale structure within a unified framework.

The absence of a statistically significant trend in the global annual mean SCM depth should not be interpreted as evidence that the underlying physical controls remained unchanged. Observational evidence indicates that ocean stratification has strengthened over recent decades (Cheng et al., 2025), while regional changes in stratification, mixing, nutrient resupply, and seasonal structure may produce heterogeneous SCM-depth responses that are not captured by a single global mean trend. The deepening tendencies observed in some oligotrophic subtropical gyre regions may be associated with stronger stratification and a deeper nutrient-resupply horizon, whereas shoaling tendencies in selected equatorial and high-latitude regions may reflect enhanced mixing, increased nutrient fluxes, or seasonal structural reorganization (Cornec et al., 2021; Xing et al., 2023; Yasunaka et al., 2022). These explanations should be regarded as plausible mechanisms rather than direct attribution, because the present analysis does not explicitly diagnose the underlying physical drivers.

These conclusions should nevertheless be interpreted within the applicability range of the dataset. Because the present method is fundamentally a statistical reconstruction constrained by observational samples, its reliability depends on the coverage of the training data across regions, seasons, and depth conditions. Here, regions with sparse observational support refer primarily to areas where the retained BGC-Argo Chl *a* profiles used for model development are sparse or patchily distributed, including parts of the high-latitude and ice-influenced oceans, tropical and subtropical gyre interiors, marginal seas, coastal-boundary regions, and other areas with limited biogeochemical-profile support (Fig. B1b). During generation of the gridded product, data-sparse regions additionally refer to areas with limited Core-Argo inference-profile support, where the gridded field may depend more strongly on larger adaptive radii or nearest-neighbor filling (Fig. B2 and Appendix B). Representativeness and interpolation-related uncertainties may therefore be larger in these regions. In addition, reconstruction errors remain relatively concentrated near the Chl *a* maximum layer because of strong vertical gradients and complex controlling mechanisms. Accordingly, this dataset is best suited for monthly mean, climatological, and basin- to global-scale analyses of spatiotemporal variability, vertical ecological structure, and model evaluation. Applications to local, event-scale, or small-scale frontal processes should be interpreted together with original observations, regional datasets, or high-resolution model outputs.

Although an adaptive-radius interpolation strategy constrained by minimum neighbor counts was adopted to balance local representativeness and spatial coverage, the reconstructed field in extremely data-sparse regions inevitably reflects regional background conditions more strongly than local variability.

## 7 Code and data availability

The global 3D Chl *a* dataset developed in this study is publicly available in a repository for long-term access. To ensure seamless compatibility across diverse analytical platforms and software environments, the dataset is provided in NetCDF format, strictly adhering to standard climate and oceanographic data conventions. The dataset is publicly available from Zenodo (Meng et al., 2026a) at https://doi.org/10.5281/zenodo.22726410. It is provided as a single file containing global monthly mean 3D Chl *a* fields for the period January 2005–June 2025. The product is gridded at a horizontal spatial resolution of $1° \times 1°$ and uses a uniform set of standardized vertical depth levels. The file contains complete temporal, spatial, and depth dimensions.

To ensure transparency and reproducibility, the core source code for the data-driven reconstruction framework is also publicly available from Zenodo (Meng et al., 2026b) at https://doi.org/10.5281/zenodo.22734381. The repository includes the PST model architecture together with the associated training and inference scripts.

## 8 Conclusions

Based on the global Argo observing system, this study constructs a global 3D Chl *a* dataset covering 2005–2025. By combining BGC-Argo Chl *a* profiles with Core-Argo temperature-salinity profiles, satellite-derived surface Chl *a* information, and a multimodal Transformer model, this study reconstructs vertically resolved Chl *a* structure for large numbers of Core-Argo profiles and extends discrete profile observations into a continuous gridded data product suitable for long-term variability analysis.

Model evaluation results show that the proposed framework maintains good predictive skill under random, year-based, and spatially independent partitioning strategies, and also shows reasonable agreement with independent ship-based observations. The reconstructed product reproduces

the global month-depth structure of Chl *a*, latitudinal differences in vertical distribution, and the major spatial patterns of the diagnosed upper-ocean Chl *a* maximum depth, indicating that the product not only maintains quantitative agreement with observations but also captures the known large-scale vertical organization of Chl *a*.

Further analyses based on this dataset indicate that SCM-depth variability during 2005–2025 is mainly characterized by pronounced regional heterogeneity and a persistent latitudinal background structure; in contrast, the global annual mean SCM depth does not show a statistically significant long-term deepening or shoaling trend. This result highlights the value of the reconstructed 3D product for examining regional SCM-depth variability and large-scale vertical ecological structure within a unified global framework. Overall, this study provides a long-term, observation-constrained global 3D Chl *a* data product that supports analyses of Chl *a* vertical structure, SCM-depth variability, and their regional differences in the global ocean.

## Appendix A: Additional model formulation and sensitivity analysis

### A1 Approximate density-anomaly-related input feature

The auxiliary density-anomaly-related feature used as a profile input was calculated from temperature $T$ and practical salinity $S$. The resulting density-anomaly-related feature was subsequently concatenated with the other profile predictors. The approximate density anomaly was defined as

$$\sigma_\rho = \rho_{\mathrm{simp}}(T, S) - 1000. \tag{A1}$$

The simplified, pressure-independent density term was calculated as

$$\rho_{\mathrm{simp}}(T, S) = \rho_{\mathrm{w}}(T) + A(T)\,S + B(T)\,S^{1.5}, \tag{A2}$$

where

$$\begin{aligned}\rho_{\mathrm{w}}(T) = {} & 999.842594 + 6.793952 \times 10^{-2} T \\ & - 9.09529 \times 10^{-3} T^2 + 1.001685 \times 10^{-4} T^3 \\ & - 1.120083 \times 10^{-6} T^4 + 6.536332 \times 10^{-9} T^5,\end{aligned} \tag{A3}$$

$$\begin{aligned}A(T) = {} & 8.24493 \times 10^{-1} - 4.0899 \times 10^{-3} T + 7.6438 \\ & \times 10^{-5} T^2 - 8.2467 \times 10^{-7} T^3 + 5.3875 \times 10^{-9} T^4,\end{aligned} \tag{A4}$$

and

$$\begin{aligned}B(T) = {} & -5.72466 \times 10^{-3} + 1.0227 \times 10^{-4} T \\ & - 1.6546 \times 10^{-6} T^2.\end{aligned} \tag{A5}$$

Here, $T$ is expressed in degrees Celsius, $S$ denotes practical salinity, and the resulting density-related quantities are expressed in $\mathrm{kg\,m^{-3}}$. This feature was used only as an approximate, pressure-independent predictor of water-mass density structure and should not be interpreted as a TEOS-10 potential-density estimate.

### A2 Sensitivity to the prescribed depth-gate decay scale

The PST model uses a prescribed depth-dependent gate to attenuate the contribution of satellite-derived surface information with increasing depth:

$$g(z_k) = \exp\left(-\frac{z_k}{\tau}\right), \tag{A6}$$

where $z_k = (k-1)/(L-1)$ is the normalized depth index and $\tau$ is the prescribed decay scale. The baseline reconstruction used $\tau = 0.3$.

To evaluate whether reconstruction performance depended strongly on this prescribed value, additional experiments were performed using $\tau = 0.2$, 0.3, 0.5, and 1.0, together with a configuration in which the depth gate was removed. All configurations were evaluated using the same validation samples and model-evaluation procedure.

Across all tested settings, RMSE varied only from 0.2250 to 0.2254, MAE from 0.1580 to 0.1585, and $R^2$ from 0.8919 to 0.8922. These small differences indicate that the reconstruction was not highly sensitive to the exact prescribed value of $\tau$. The baseline value of $\tau = 0.3$ was therefore retained because it provides a gradual attenuation of surface satellite information with depth while maintaining stable reconstruction performance.

**Table A1.** Sensitivity of reconstruction performance to the prescribed depth-gate decay scale.

| Experiment | $N$ | RMSE | MAE | Bias | $R$ | $R^2$ |
|---|---|---|---|---|---|---|
| no_gate | 816 815 | 0.2254 | 0.1585 | −0.0057 | 0.9444 | 0.8919 |
| $\tau = 0.2$ | 816 815 | 0.2252 | 0.1582 | −0.0047 | 0.9445 | 0.8920 |
| $\tau = 0.3$ (baseline) | 816 815 | 0.2250 | 0.1580 | −0.0089 | 0.9447 | 0.8922 |
| $\tau = 0.5$ | 816 815 | 0.2253 | 0.1584 | −0.0058 | 0.9445 | 0.8919 |
| $\tau = 1.0$ | 816 815 | 0.2253 | 0.1583 | −0.0106 | 0.9446 | 0.8919 |

$N$ denotes the number of valid prediction-observation pairs. RMSE, MAE, and bias were calculated in $\log_{10}$-transformed Chl *a* space; $R$ and $R^2$ are dimensionless.

## Appendix B: Sampling support and interpolation diagnostics

### B1 Spatial and temporal distributions of the Argo profiles

The reconstruction framework transfers information learned from the relatively sparse BGC-Argo Chl *a* profiles to the substantially larger Core-Argo temperature-salinity profile network. Figure B1 compares the spatial and temporal distributions of the two profile datasets actually used in the reconstruction workflow.

The BGC-Argo distribution corresponds to the 109 083 profiles retained for model training, validation, and testing after quality screening. The Core-Argo distribution corresponds to the 2 816 143 valid profiles used for large-scale Chl *a* inference. The comparison illustrates the broader spatial and temporal coverage of Core-Argo and identifies regions where direct BGC-Argo observational support remains relatively sparse.

### B2 Random holdout evaluation of the maximum adaptive interpolation radius

The final gridded product was generated using an adaptive-radius Gaussian-weighted interpolation scheme. Candidate radii of 100, 200, 300, 500, and 700 km were considered, and the smallest radius satisfying the prescribed neighbor requirement was selected for each target grid cell.

To evaluate the influence of the maximum candidate radius, a random subset of the inferred Chl *a* profiles was withheld from the interpolation. The remaining inferred profiles were then used to reconstruct Chl *a* at the withheld locations. All configurations used the same adaptive interpolation logic, with only the maximum permitted search radius being changed.

Coverage increased substantially as the maximum radius increased, whereas the interpolation errors changed only slightly. The 700 km setting increased adaptive-stage coverage from 99.57 % to 99.91 % relative to the 500 km setting, while RMSE increased from 0.02359 to 0.02373 mg $m^{-3}$. Therefore, 700 km was not treated as the nominal interpolation radius or as a universally optimal local scale. It was retained only as an upper-bound fallback that improves coverage in regions where smaller radii do not contain sufficient neighboring profiles.

### B3 Actual usage of the adaptive interpolation radii

The actual radius selected at each ocean grid cell depended on the number of nearby inferred profiles. Smaller radii were used wherever sufficient local profile support was available. Table B2 summarizes the frequency with which each radius was selected during the adaptive interpolation stage.

The 300 and 500 km radii together accounted for 62.41 % of successful adaptive interpolation cases, whereas the 700 km radius accounted for 12.12 %. The spatial contribution of 700 km radius selections was not uniformly distributed over the global ocean but was concentrated mainly in high-latitude regions, marginal seas, coastal-boundary areas, and other sparsely sampled regions (Fig. B2). These results indicate that 700 km was used as a limited fallback radius in regions with weak local profile support rather than as the dominant interpolation scale of the product.

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 gray.

### B4 Ancillary sampling-support and interpolation diagnostics

In addition to the final gridded Chl *a* field, the updated NetCDF product includes ancillary variables describing sampling support and interpolation dependence. At the time-depth level, n_points_used and n_profiles_used record the numbers of inferred Chl *a* points and profiles available for each month and depth level; nan_count_global and nan_count_ocean record the numbers of missing grid cells over the complete grid and within the ocean domain, respectively; ocean_coverage_after_nearest records the fraction of ocean grid cells containing valid Chl *a* after nearest-neighbor filling; and adaptive_fail_counts_tz, nearest_fill_counts_tz, and final_fail_counts_tz record the numbers of adaptive-

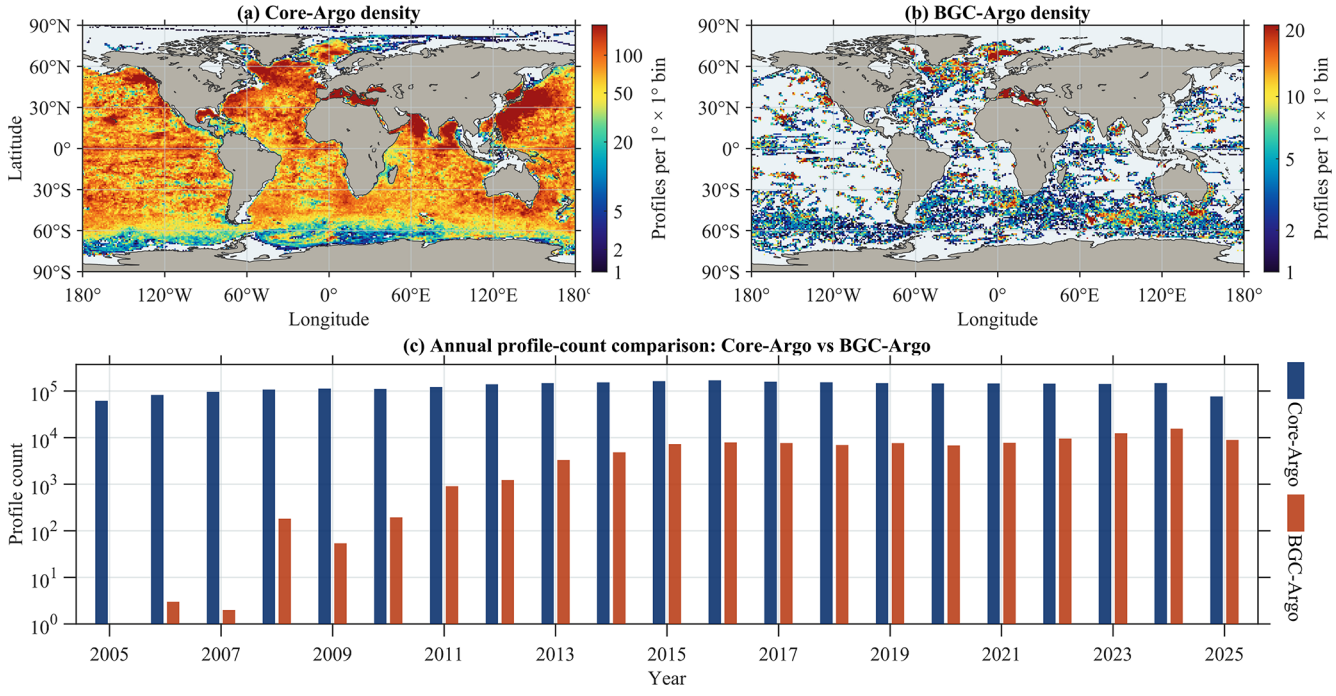



**Figure B1.** Distribution comparison of Core-Argo inference profiles and BGC-Argo Chl *a* profiles used for model development. **(a)** Full-period spatial density of the Core-Argo profiles used for inference, aggregated on a 1° grid. **(b)** Full-period spatial density of the BGC-Argo Chl *a* profiles retained for model training, validation, and testing, aggregated on a 1° grid. Colors in panels **(a)** and **(b)** are mapped using $\log_{10}(N+1)$, with color-bar tick labels showing profile counts ($N$). **(c)** Annual numbers of Core-Argo inference profiles and retained BGC-Argo Chl *a* profiles. Profile counts are shown on a logarithmic scale to facilitate comparison between the two datasets.

**Table B1.** Random holdout evaluation for different maximum adaptive interpolation radii.

| 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 |

Coverage represents the fraction of withheld inferred 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-neighbor filling step was not included in this holdout evaluation; therefore, the table evaluates the adaptive-radius interpolation stage itself rather than the complete model-plus-interpolation error against independent observations.

interpolation failures, nearest-filled cells, and cells remaining missing after the complete interpolation procedure.

The variables r_bins_km and r_usage_counts provide the candidate radius categories and their corresponding numbers of successful selections. At the grid-cell level, interp_radius_km records the selected adaptive interpolation radius, interp_neighbor_count records the number of neighboring inferred profiles supporting each estimate, adaptive_fail_before_nearest identifies cells not reconstructed during adaptive interpolation, nearest_fill_used identifies cells subsequently completed by nearest-neighbor filling, and nearest_fill_distance_km records the distance to the source cell used for filling.

These variables allow users to identify regions, months, and depth levels with different levels of profile support and interpolation dependence. They are intended as practical reliability diagnostics rather than as a formal decomposition of observational, model, representativeness, and interpolation uncertainties.

**Table B2.** Actual usage of the 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 % |

Radius usage was calculated only for grid cells successfully reconstructed during the adaptive-radius stage. Cases completed subsequently by nearest-neighbor filling were not counted as adaptive-radius usage.

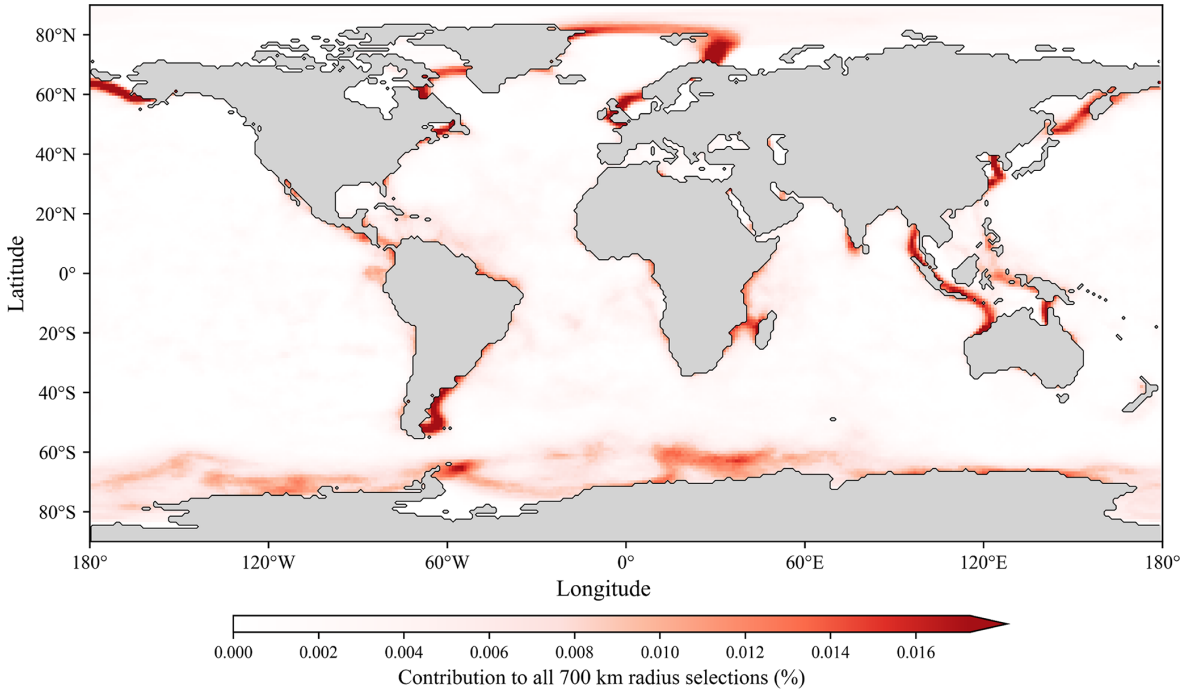



**Figure B2.** Spatial distribution of the contribution to the use of the 700 km adaptive interpolation radius.

## Appendix C: Additional validation of inferred Chl *a* profiles

### C1 Statistical significance of depth-wise Spearman correlations

Spearman's $\rho$ was calculated for each 20 m depth bin by pooling all valid reconstructed-observed pairs within the corresponding interval. The resulting $p$ values were corrected across all evaluated depth bins using the Benjamini-Hochberg false discovery rate procedure. All depth bins remained statistically significant after FDR correction, with adjusted $p < 0.001$. The corresponding Spearman's $\rho$ values, ranging from 0.66 to 0.89, are shown in Fig. 3a5.

### C2 Profile-wise relative errors in the original linear Chl *a* domain

To complement the validation metrics presented in the main text, relative errors were additionally evaluated in the original linear Chl *a* concentration domain. For consistency with the profile-wise RMSE map, the mean absolute percentage error was first calculated across all valid depth levels within each profile, and the median was then obtained across the 16 479 validation profiles. Depth-specific values were calculated using the same procedure within each prescribed depth range.

The overall median profile-wise MAPE was 32.50 %, with depth-specific values ranging from 24.23 % to 38.07 % (Table C2). The largest value occurred at 100–200 m, coinciding

**Table C1.** Summary of the statistical significance of depth-wise Spearman correlations.

| Evaluated depth range | Depth-bin interval | Range of Spearman's $\rho$ | FDR-adjusted ($p$) value |
|---|---|---|---|
| 10–500 m | 20 m | 0.66–0.89 | ($p < 0.001$) for all bins |

Spearman's $\rho$ was calculated separately for each 20 m depth bin. The $p$ values were corrected across all evaluated bins using the Benjamini-Hochberg false discovery rate procedure.

**Table C2.** Median profile-wise MAPE in the original linear Chl *a* concentration domain.

| Depth range (m) | Median profile-wise MAPE (%) |
|---|---|
| 10–50 | 28.01 |
| 50–100 | 29.64 |
| 100–200 | 38.07 |
| 200–300 | 28.46 |
| 300–500 | 24.23 |
| 10–500 | 32.50 |

Predictions and observations were converted from $\log_{10}$-transformed values back to the original linear Chl *a* concentration domain. MAPE was first calculated across all valid depth levels within each profile, and the table reports the median across the 16 479 validation profiles.

with the depth range in which subsurface Chl *a* maxima and strong vertical gradients are commonly observed.

Because percentage errors use the observed concentration as the denominator, very low Chl *a* values can produce disproportionately large relative errors even when the corresponding absolute concentration differences are small. Given the broad dynamic range of Chl *a* and the substantial occurrence of very low concentrations, these results are provided only as supplementary indicators of profile-scale relative discrepancies. The $\log_{10}$-domain RMSE and MAE, together with $R^2$ and Spearman's $\rho$ reported in the main text, remain the primary metrics for evaluating the reconstruction of the broad dynamic range and large-scale spatial and vertical variability of Chl *a*.

**Video supplement.** A visualization of the global 3D chlorophyll *a* dataset is available at https://doi.org/10.5281/zenodo.22808725 (Meng et al., 2026c).

**Author contributions.** Yuntao Wang conceived the study, directed the research, and supervised the experimental process. Ziyue Meng developed the methodology, performed the data analysis and model training, and wrote the original manuscript draft. Lei Han assisted in project supervision. All authors contributed to the critical review and editing of the manuscript.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

**Acknowledgements.** The authors would like to express their sincere gratitude to the Hangzhou Global Ocean Argo System Observation and Research Station (http://www.argo.org.cn/, last access: 27 September 2026) for providing the essential Argo profile data used in this study. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it (https://argo.ucsd.edu, last access: 27 September 2026; https://www.ocean-ops.org, last access: 27 September 2026). The Argo Program is part of the Global Ocean Observing System. We also acknowledge the NASA Ocean Biology Processing Group (OBPG) for providing the MODIS data (https://oceancolor.gsfc.nasa.gov/, last access: 27 September 2026).

**Financial support.** This research has been supported by the National Key Research and Development Program of China (grant no. 2023YFC2811800), National Natural Science Foundation of China (grant nos. 4257061617 and 42275003) and Zhejiang Provincial Natural Science Foundation of China (grant no. LR25D060001).

**Review statement.** This paper was edited by Xingchen (Tony) Wang and reviewed by two anonymous referees.

## References

Argo Data Management Team: Argo user's manual, version 3.44.0, https://doi.org/10.13155/29825, 2025.

Arteaga, L. A. and Rousseaux, C. S.: Evaluation of vertical patterns in chlorophyll-*a* derived from a data-assimilating model of satellite-based ocean color, Earth Space Sci., 11, e2023EA003378, https://doi.org/10.1029/2023EA003378, 2024.

Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., Schmechtig, C., Taillandier,

V., Leymarie, E., Penkerc'h, C., D'Ortenzio, F., Claustre, H., and Bricaud, A.: Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database, Biogeosciences, 16, 1321–1342, https://doi.org/10.5194/bg-16-1321-2019, 2019.

Barnes, S. L.: A technique for maximizing details in numerical weather-map analysis, J. Appl. Meteorol., 3, 396–409, https://doi.org/10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2, 1964.

Beckers, J.-M. and Rixen, M.: EOF calculations and data filling from incomplete oceanographic data sets, J. Atmos. Ocean. Tech., 20, 1839–1856, https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2, 2003.

Behrenfeld, M. J., Marañón, E., Siegel, D. A., and Hooker, S. B.: Photoacclimation and nutrient based model of light-saturated photosynthesis for quantifying oceanic primary production, Mar. Ecol. Prog. Ser., 228, 103–117, https://doi.org/10.3354/meps228103, 2002.

Bouman, H. A., Jackson, T., Sathyendranath, S., and Platt, T.: Vertical structure in chlorophyll profiles: influence on primary production in the Arctic Ocean, Philos. T. R. Soc. A, 378, 20190351, https://doi.org/10.1098/rsta.2019.0351, 2020.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S.: End-to-end object detection with transformers, in: Computer Vision – ECCV 2020, Springer, Cham, 213–229, https://doi.org/10.1007/978-3-030-58452-8_13, 2020.

Chen, J., Gong, X., Guo, X., Xing, X., Lu, K., Gao, H., and Gong, X.: Improved perceptron of subsurface chlorophyll maxima by a deep neural network: a case study with BGC-Argo float data in the northwestern Pacific Ocean, Remote Sens., 14, 632, https://doi.org/10.3390/rs14030632, 2022.

Cheng, L., Li, G., Long, S.-M., Li, Y., von Schuckmann, K., Trenberth, K. E., Mann, M. E., Abraham, J., Du, Y., Cheng, X., Liu, H., Xu, Z., Liu, M., Peng, Q., Gong, X., Ma, Z., and Yuan, H.: Ocean stratification in a warming climate, Nat. Rev. Earth Environ., 6, 637–655, https://doi.org/10.1038/s43017-025-00715-5, 2025.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the Global Ocean with Biogeochemical-Argo, Annu. Rev. Mar. Sci., 12, 23–48, https://doi.org/10.1146/annurev-marine-010419-010956, 2020.

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., D'Ortenzio, F., Gentili, B., and Schmechtig, C.: Deep chlorophyll maxima in the global ocean: occurrences, drivers and characteristics, Global Biogeochem. Cy., 35, e2020GB006759, https://doi.org/10.1029/2020GB006759, 2021.

Cullen, J. J.: Subsurface chlorophyll maximum layers: enduring enigma or mystery solved?, Annu. Rev. Mar. Sci., 7, 207–239, https://doi.org/10.1146/annurev-marine-010213-135111, 2015.

Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, Earth Syst. Sci. Data, 6, 273–284, https://doi.org/10.5194/essd-6-273-2014, 2014.

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, Nat. Rev. Methods Primers, 2, 76, https://doi.org/10.1038/s43586-022-00154-2, 2022.

Fofonoff, N. P. and Millard Jr., R. C.: Algorithms for computation of fundamental properties of seawater, UNESCO Technical Papers in Marine Science, No. 44, UNESCO, Paris, 53 pp., https://www.jodc.go.jp/info/ioc_doc/UNESCO_tech/059832eb.pdf (last access: 27 September 2026), 1983.

Glover, D. M., Doney, S. C., Oestreich, W. K., and Tullo, A. W.: Geostatistical analysis of mesoscale spatial variability and error in SeaWiFS and MODIS/Aqua global ocean color data, J. Geophys. Res.-Oceans, 123, 22–39, https://doi.org/10.1002/2017JC013023, 2018.

Hu, Q., Chen, X., Bai, Y., He, X., Li, T., and Pan, D.: Reconstruction of 3-D ocean chlorophyll-*a* structure in the northern Indian Ocean using satellite and BGC-Argo data, IEEE T. Geosci. Remote, 61, 4200513, https://doi.org/10.1109/TGRS.2022.3233385, 2023.

Jackson, T., Calton, B., and Hockley, K.: C3S ocean colour version 6.0: product user guide and specification, issue 1.1, Copernicus Climate Change Service, Reading, UK, https://dast.copernicus-climate.eu/documents/satellite-ocean-colour/v6.0/WP2-FDDP-2022-04_C3S2-Lot3_PUGS-of-v6.0-OceanColour-product_v1.1_FINAL.pdf (last access: 27 September 2026), 2023.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P.: On the representation error in data assimilation, Q. J. Roy. Meteor. Soc., 144, 1257–1278, https://doi.org/10.1002/qj.3130, 2018.

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array, J. Geophys. Res.-Oceans, 122, 6416–6436, https://doi.org/10.1002/2017JC012838, 2017.

Koldasbayeva, D. and Zaytsev, A.: Foundation for unbiased cross-validation of spatio-temporal models for species distribution modeling, Ecol. Inform., 92, 103521, https://doi.org/10.1016/j.ecoinf.2025.103521, 2025.

Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B. F., Bouman, H. A., Babin, M., Brewin, R. J. W., Doblin, M., Estrada, M., Figueiras, F. G., Furuya, K., González-Benítez, N., Gudfinnsson, H. G., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V. A., Marañón, E., Raman, M., Richardson, K., Rozema, P. D., van de Poll, W. H., Segura, V., Tilstone, G. H., Uitz, J., van Dongen-Vogels, V., Yoshikawa, T., and Sathyendranath, S.: Primary production, an index of climate change in the ocean: satellite-based estimates over two decades, Remote Sens., 12, 826, https://doi.org/10.3390/rs12050826, 2020.

Li, H., Xu, F., Zhou, W., Wang, D., Wright, J. S., Liu, Z., and Lin, Y.: Development of a global gridded Argo data set with Barnes successive corrections, J. Geophys. Res.-Oceans, 122, 866–889, https://doi.org/10.1002/2016JC012285, 2017.

Li, J. and Heap, A. D.: A review of spatial interpolation methods for environmental scientists, Geoscience Australia, Record 2008/23, Canberra, https://www.ga.gov.au/bigobj/GA12526.pdf (last access: 27 September 2026), 2008.

Liu, X. and Wang, M.: Gap filling of missing data for VIIRS global ocean color products using the DINEOF method, IEEE T. Geosci. Remote, 56, 4464–4476, https://doi.org/10.1109/TGRS.2018.2820423, 2018.

Lomas, M. W. and Bates, N. R.: Potential controls on interannual partitioning of organic carbon during the winter/spring phytoplankton bloom at the Bermuda Atlantic Time-series

Study (BATS) site, Deep-Sea Res. Pt. I, 51, 1619–1636, https://doi.org/10.1016/j.dsr.2004.06.007, 2004.

Mamnun, N., Völker, C., Vrekoussis, M., and Nerger, L.: Uncertainties in ocean biogeochemical simulations: application of ensemble data assimilation to a one-dimensional model, Front. Mar. Sci., 9, 984236, https://doi.org/10.3389/fmars.2022.984236, 2022.

Marañón, E., Van Wambeke, F., Uitz, J., Boss, E. S., Dimier, C., Dinasquet, J., Engel, A., Haëntjens, N., Pérez-Lorenzo, M., Taillandier, V., and Zäncker, B.: Deep maxima of phytoplankton biomass, primary production and bacterial production in the Mediterranean Sea, Biogeosciences, 18, 1749–1767, https://doi.org/10.5194/bg-18-1749-2021, 2021.

Martinez, E., Gorgues, T., Lengaigne, M., Sauzède, R., Menkes, C., Uitz, J., Di Lorenzo, E., and Fablet, R.: Reconstructing global chlorophyll-*a* variations using a non-linear statistical approach, Front. Mar. Sci., 7, 464, https://doi.org/10.3389/fmars.2020.00464, 2020.

Masuda, Y., Yamanaka, Y., Smith, S. L., Hirata, T., Nakano, H., Oka, A., and Sumata, H.: Photoacclimation by phytoplankton determines the distribution of global subsurface chlorophyll maxima in the ocean, Commun. Earth Environ., 2, 128, https://doi.org/10.1038/s43247-021-00201-y, 2021.

Mattei, F. and Scardi, M.: Collection and analysis of a global marine phytoplankton primary-production dataset, Earth Syst. Sci. Data, 13, 4967–4985, https://doi.org/10.5194/essd-13-4967-2021, 2021.

McDougall, T. J. and Barker, P. M.: Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox, SCOR/IAPSO WG127, 28 pp., ISBN 978-0-646-55621-5, 2011.

Meng, Z., Wang, Y., and Han, L.: Global 3D Chlorophyll-*a* Dataset Derived from Multimodal Deep Learning Reconstruction (2005–2025), Zenodo [data set], https://doi.org/10.5281/zenodo.22726410, 2026a.

Meng, Z., Wang, Y., and Han, L.: PST_Global_3D_Chl_Reconstruction_Code, Zenodo [code], https://doi.org/10.5281/zenodo.22734381, 2026b.

Meng, Z., Wang, Y., and Han, L.: Visualization of the Global 3D Chlorophyll *a* Dataset, Zenodo [video], https://doi.org/10.5281/zenodo.22808725, 2026c.

Mignot, A., Claustre, H., Uitz, J., Poteau, A., D'Ortenzio, F., and Xing, X.: Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: a Bio-Argo float investigation, Global Biogeochem. Cy., 28, 856–876, https://doi.org/10.1002/2013GB004781, 2014.

Millero, F. J. and Poisson, A.: International one-atmosphere equation of state of seawater, Deep-Sea Res. Pt. A, 28, 625–629, https://doi.org/10.1016/0198-0149(81)90122-9, 1981.

O'Reilly, J. E. and Werdell, P. J.: Chlorophyll algorithms for ocean color sensors – OC4, OC5 and OC6, Remote Sens. Environ., 229, 32–47, https://doi.org/10.1016/j.rse.2019.04.021, 2019.

Park, J., Kim, J.-H., Kim, H.-C., Kim, B.-K., Bae, D., Jo, Y.-H., Jo, N., and Lee, S. H.: Reconstruction of ocean color data using machine learning techniques in polar regions: focusing on off Cape Hallett, Ross Sea, Remote Sens., 11, 1366, https://doi.org/10.3390/rs11111366, 2019.

Pottier, C., Garçon, V., Larnicol, G., Sudre, J., Schaeffer, P., and Le Traon, P.-Y.: Merging SeaWiFS and MODIS/Aqua ocean color data in north and equatorial Atlantic using weighted averaging and objective analysis, IEEE T. Geosci. Remote, 44, 3436–3451, https://doi.org/10.1109/TGRS.2006.878441, 2006.

Richardson, A. J., Pfaff, M. C., Field, J. G., Silulwane, N. F., and Shillington, F. A.: Identifying characteristic chlorophyll *a* profiles in the coastal domain using an artificial neural network, J. Plankton Res., 24, 1289–1303, https://doi.org/10.1093/plankt/24.12.1289, 2002.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H. C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D. Ó., d'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Le Traon, P.-Y., Lenn, Y.-D., Maze, G., Mork, K. A., Morris, T., Nagai, T., Nash, J., Naveira Garabato, A., Olsen, A., Pattabhi, R. R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T. W., Turton, J., Velez-Belchi, P. J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A. F., Waterman, S., Watson, A., Wilson, C., Wong, A. P. S., Xu, J., and Yasuda, I.: On the future of Argo: a global, full-depth, multi-disciplinary array, Front. Mar. Sci., 6, 439, https://doi.org/10.3389/fmars.2019.00439, 2019.

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., D'Ortenzio, F., Ras, J., Drapeau, S., Haëntjens, N., and Barbieux, M.: Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: a global analysis of WET Labs ECO sensors, Limnol. Oceanogr.-Meth., 15, 572–585, https://doi.org/10.1002/lom3.10185, 2017.

Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., and Martinez, E.: A multi-mode convolutional neural network to reconstruct satellite-derived chlorophyll-*a* time series in the global ocean from physical drivers, Front. Mar. Sci., 10, 1077623, https://doi.org/10.3389/fmars.2023.1077623, 2023.

Sathyendranath, S., Brewin, R. J. W., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A. B., Dingle, J., Doerffer, R., Donlon, C., Dowell, M., Farman, A., Grant, M., Groom, S., Horseman, A., Jackson, T., Krasemann, H., Lavender, S., Martinez-Vicente, V., Mazeran, C., Mélin, F., Moore, T. S., Müller, D., Regner, P., Roy, S., Steele, C. J., Steinmetz, F., Swinton, J., Taberner, M., Thompson, A., Valente, A., Zühlke, M., Brando, V. E., Feng, H., Feldman, G., Franz, B. A., Frouin, R., Gould, R. W., Jr., Hooker, S. B., Kahru, M., Kratzer, S., Mitchell, B. G., Muller-Karger, F. E., Sosik, H. M., Voss, K. J., Werdell, J., and Platt, T.: An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI), Sensors, 19, 4285, https://doi.org/10.3390/s19194285, 2019.

Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall'Olmo, G., D'Ortenzio, F., Gentili, B., Poteau, A., and Schmechtig, C.: A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: retrieval of the particulate backscatter-

ing coefficient, J. Geophys. Res.-Oceans, 121, 2552–2571, https://doi.org/10.1002/2015JC011408, 2016.

Schmechtig, C., Claustre, H., Poteau, A., D'Ortenzio, F., Schallenberg, C., Trull, T. W., and Xing, X.: Biogeochemical-Argo quality control manual for chlorophyll-*a* concentration and chl-fluorescence, Version 3.0, Ifremer, https://doi.org/10.13155/35385, 2023.

Schollaert Uz, S., Busalacchi, A. J., Smith, T. M., Evans, M. N., Brown, C. W., and Hackert, E. C.: Interannual and decadal variability in tropical Pacific chlorophyll from a statistical reconstruction: 1958–2008, J. Climate, 30, 7293–7315, https://doi.org/10.1175/JCLI-D-16-0202.1, 2017.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J. G., Li, H., Long, M. C., Luo, J. Y., Nakano, H., Romanou, A., Schwinger, J., Stock, C., Santana-Falcón, Y., Takano, Y., Tjiputra, J., Tsujino, H., Watanabe, M., Wu, T., Wu, F., and Yamamoto, A.: Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6, Curr. Clim. Change Rep., 6, 95–119, https://doi.org/10.1007/s40641-020-00160-0, 2020.

Siegel, D. A., DeVries, T., Cetinić, I., and Bisson, K. M.: Quantifying the ocean's biological pump and its carbon cycle impacts on global scales, Annu. Rev. Mar. Sci., 15, 329–356, https://doi.org/10.1146/annurev-marine-040722-115226, 2023.

Silsbe, G. M., Fox, J., Westberry, T. K., and Halsey, K. H.: Global declines in net primary production in the ocean color era, Nat. Commun., 16, 5821, https://doi.org/10.1038/s41467-025-60906-y, 2025.

Stock, A.: Choosing blocks for spatial cross-validation: lessons from a marine remote sensing case study, Front. Remote Sens., 6, 1531097, https://doi.org/10.3389/frsen.2025.1531097, 2025.

Udaya Bhaskar, T. V. S., Ravichandran, M., and Devender, R.: An operational objective analysis system at INCOIS for generation of Argo value added products, Technical Report No. INCOIS/MOG-TR-2/07, Indian National Centre for Ocean Information Services, Hyderabad, India, https://frouingroup.ucsd.edu/ddeell/Incois-OA.pdf (last access: 27 September 2026), 2007.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Adv. Neur. In., 30, 5998–6008, 2017.

Wojtasiewicz, B., Hardman-Mountford, N. J., Antoine, D., Dufois, F., Slawinski, D., and Trull, T. W.: Use of bio-optical profiling float data in validation of ocean colour satellite products in a remote ocean region, Remote Sens. Environ., 209, 275–290, https://doi.org/10.1016/j.rse.2018.02.057, 2018.

Xing, X., Morel, A., Claustre, H., Antoine, D., D'Ortenzio, F., Poteau, A., and Mignot, A.: Combined processing and mutual interpretation of radiometry and fluorimetry from autonomous profiling Bio-Argo floats: chlorophyll-*a* retrieval, J. Geophys. Res., 116, C06020, https://doi.org/10.1029/2010JC006899, 2011.

Xing, X., Xiu, P., Laws, E. A., Yang, G., Liu, X., and Chai, F.: Light-driven and nutrient-driven displacements of subsurface chlorophyll maximum depth in subtropical gyres, Geophys. Res. Lett., 50, e2023GL104510, https://doi.org/10.1029/2023GL104510, 2023.

Yang, K., Meyer, A., Strutton, P. G., and Fischer, A. M.: Global trends of fronts and chlorophyll in a warming ocean, Commun. Earth Environ., 4, 489, https://doi.org/10.1038/s43247-023-01160-2, 2023.

Yang, Y., Li, X., and Li, X.: A self-attention-based deep learning model for estimating global phytoplankton pigment profiles, IEEE T. Geosci. Remote, 62, 4208415, https://doi.org/10.1109/TGRS.2024.3435044, 2024.

Yasunaka, S., Ono, T., Sasaoka, K., and Sato, K.: Global distribution and variability of subsurface chlorophyll *a* concentrations, Ocean Sci., 18, 255–268, https://doi.org/10.5194/os-18-255-2022, 2022.

Zhan, H., Shi, P., and Chen, C.: Retrieval of oceanic chlorophyll concentration using support vector machines, IEEE T. Geosci. Remote, 41, 2947–2951, https://doi.org/10.1109/TGRS.2003.819870, 2003.

Zhao, X., Gong, X., Gong, X., Liu, J., Wang, G., Wang, L., Guo, X., and Gao, H.: Evolution of 3-D chlorophyll in the northwestern Pacific Ocean using a Gaussian-activation deep neural network model, Front. Mar. Sci., 11, 1378488, https://doi.org/10.3389/fmars.2024.1378488, 2024.

Zürcher, B. K.: Fast approximate Barnes interpolation: illustrated by Python-Numba implementation fast-barnes-py v1.0, Geosci. Model Dev., 16, 1697–1711, https://doi.org/10.5194/gmd-16-1697-2023, 2023.