https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

The first consistent global gap-free daily 4 km dataset of particulate organic carbon, particulate organic nitrogen and their ratio (1998–2023)

Yu Zhang[1,2,3,4,5], Huizeng Liu[1,2,3,4,5*], Cong Liu[6], Nan Xu[2,3,4], Lin Yan[1,2,3,4], Chao Yang[2,3,4], Yongquan Wang[1,2,3,4], Guofeng Wu[2,3,4], Qingquan Li[1,2,3,4]

[1]Institute for Advanced Study & Tiandu-Shenzhen University Deep Space Exploration Joint Laboratory & Space Science Center, Shenzhen University, Shenzhen,·518060, China
[2]MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen, 518060, China
[3]Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen, 518060, China
[4] Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen, 518060, China
[5]Institute for Carbon Neutrality, Shenzhen University, Shenzhen, 518060, China
[6]State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, 510301, China

Correspondence to: Huizeng Liu (huizeng.liu@szu.edu.cn)

Abstract. Particulate organic carbon (POC), particulate organic nitrogen (PON) and their ratios (POC:PON) are key to understanding particulate organic matter (POM) and marine biogeochemical cycles. However, global observations remain insufficient, with existing products suffering from spatial and temporal gaps, inconsistent methodologies, and interruptions from clouds and satellite sampling limitations. Here, we present a global gap-free daily 4 km dataset of consistent POC, PON and POC:PON from January 1998 to December 2023 (GGFD-POM), generated using a concise retrieval then reconstruction workflow. POC and PON concentrations were retrieved from ocean color data (OC-CCI v6.0) and Copernicus reanalysis physics data using newly developed Gaussian Process Regression (GPR) models. The models were trained on ~3110 matchups between in-situ observations, OC-CCI bio-optical properties, and Copernicus physical properties, achieving R^2 of 0.87 and 0.89, and RMSE of 1.47 mg m^{-3} and 1.41 mg m^{-3} for POC and PON, respectively. The Discrete Cosine Transform–Penalized Least Squares (DCT-PLS) approach was subsequently applied to reconstruct missing values in the satellite-retrieved POC and PON fields, resulting in gap-free global datasets. Based on these reconstructed products, a gap-free global POC:PON dataset was further derived. Validation using an independent in-situ dataset confirmed high accuracy of both satellite retrievals and reconstructions for POC, PON and POC:PON. Triple-collocation analysis (TCA) exhibited that the GGFD POC data outperform existing MODIS-Aqua and MULTIOBS POC datasets, reflecting the distinct strengths of the GGFD-POM product. Comparative analysis of spatiotemporal variations in POC, PON, and POC:PON further demonstrated that the gap-free dataset better captures trends and magnitudes, enhancing understanding of their roles in the global carbon

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

and nitrogen cycles. The complete GGFD POM product dataset (1998–2023) is openly available at https://doi.org/10.11888/Ocean.tpdc.303488 (Zhang and Liu, 2026).

1 Introduction

The POC and PON are primarily produced by phytoplankton through photosynthesis, with additional contributions from the degradation of organic matter by microorganisms and other biological and physical processes (Aumont et al., 2017; Falkowski et al., 1998; Kharbush et al., 2020). They are key contributors to marine biogeochemical processes and carbon cycling in the ocean, thereby influencing the climate system. As these particles, sink out of the surface ocean, they transport carbon and nutrients to the deep sea, thereby facilitating the drawdown of atmospheric CO_2 (Boyd and Trull, 2007; Boyd et al., 2019). POC and PON are the principal elemental constituents of POM (Martiny et al., 2013a; Fumenia et al., 2025). Variations in the POC:PON ratios influence carbon and nutrient fluxes, the efficiency of the biological pump, as well as processes ranging from primary production to export (Demir et al., 2025; Copinmontegut and Copinmontegut, 1983; Matsumoto et al., 2020; Taucher et al., 2021).

POC and PON concentrations exhibit significant spatiotemporal variability across the global ocean, reflecting the combined influences of phytoplankton growth and decomposition together with physical transport processes, including vertical mixing and horizontal advection. In contrast, the POC:PON ratio, based on the classical Redfield ratio (C:N:P = 106:16:1) proposed by Redfield (1934), suggests that phytoplankton maintain a globally consistent elemental composition. This concept became a cornerstone of marine biogeochemistry. However, decades of research have shown that planktonic stoichiometry is flexible (Redfield et al., 1963), with no physiological mechanism to enforce its invariance across spatial and temporal scales (Geider and La Roche, 2002). Numerous basin- to global-scale studies have documented significant variations in POC:PON, such as latitudinal gradients in elemental ratios across diverse trophic regimes (Martiny et al., 2013b), and regional departures from Redfield stoichiometry (Henderson et al., 2025; Wang et al., 2025a; Zhang et al., 2024a; Fagan et al., 2024). While POC:PON variability has been increasingly documented, our understanding of both the ratio and its principal constituents, POC and PON, remains limited, primarily because globally consistent, long-term POM datasets with continuous spatial and temporal coverage are still lacking.

Traditionally, POC and PON concentrations have been measured through sparse shipboard surveys or a few time-series stations, offering limited spatial and temporal coverage and failing to capture the full heterogeneity of surface ocean biogeochemistry (Fumenia et al., 2025; Martiny et al., 2013a). During the last two decades, ocean color remote sensing has provided unprecedented opportunities for monitoring marine biogeochemical processes continuously and consistently over broad spatiotemporal scales. Among them, satellite retrieval of POC has advanced significantly, leading to three main model types: those based on inherent optical properties (IOPs), apparent optical properties (AOPs), and biological properties (Stramski et al., 2008; Le et al., 2018; Liu et al., 2019). IOPs-based models most commonly use the particulate backscattering coefficient (b_{bp}), with Stramski et al. (1999) pioneering POC estimation from b_{bp} at 510 nm. However,

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

accurately deriving IOPs from AOPs remains challenging (Jiang et al., 2019; Evers-King et al., 2017), limiting their wider application. In contrast, AOP-based models are more widely employed (Stramski et al., 2022), including the empirical algorithm used by the NASA Ocean Biology Processing Group, which relies on the blue-green remote sensing reflectance (R_{rs}) ratio (Stramski et al., 2008). These models, however, are less reliable in optically complex coastal and inland waters (Le et al., 2017; Jiang et al., 2019; Duan et al., 2014), partly due to the combined effect of phytoplankton and detrital particles on POC and taxon-dependent relationships between R_{rs} and POC (Pahlevan et al., 2017). A third category of models, based on suspended particulate matter (SPM), works effectively in turbid waters, but is limited by the spatiotemporal variability in POC:SPM ratios. In recent years, artificial intelligence (AI) techniques have been increasingly used to retrieve global oceanic POC concentrations, integrating multiple bio-optical variables into unified models for improved accuracy and robustness (Liu et al., 2021; Zhang et al., 2024d).

Compared with POC, research on satellite retrieval of PON remains relatively scarce, although interest in this topic has grown in recent years. PON concentrations have been shown to be closely associated with bio-optical properties, particularly b_{bp} and phytoplankton absorption coefficients (a_{ph}), in a wide range of aquatic environment (Fumenia et al., 2025; Fumenia et al., 2020). These relationships demonstrate the potential of bio-optical parameters for estimating PON concentrations. Wang et al. (2022b) explored polynomial regression approaches based on multiple bio-optical parameters and spectral indices, and demonstrated their applicability for remote sensing retrieval of PON. Recent application of GPR to PON retrieval across multiple ocean color satellite missions, integrating diverse bio-optical features, enhancing accuracy and generalizability (Zhang et al., 2024c). This AI-based approach, combined with NASA Level-3 POC products, enables the first global estimation of the oceanic POC:PON at high spatiotemporal resolution. In addition, initial attempts have been made to retrieve the vertical distribution of PON using AI-based remote sensing methods (Zhang et al., 2025).

While satellite retrievals of POC and PON have achieved reasonable accuracy, their global coverage is limited by the polar-orbiting configurations of ocean color sensors, which create systematic gaps between swaths. Cloud cover, sun glint effects, and unfavorable solar or viewing zenith angles can further increase the occurrence of missing observations (Liu and Wang, 2022; Wang et al., 2024; Wang et al., 2025b). These gaps hinder complete global spatiotemporal characterization. To overcome this, various data reconstruction techniques, including Data Interpolating Empirical Orthogonal Functions (DINEOF; Beckers and Rixen (2003)), data interpolating convolutional autoencoder (DINCAE; Barth et al. (2020)), and DCT-PLS (Garcia (2010)), have been used to produce continuous, gap-free datasets with preserved temporal coherence and improved spatial completeness (Liu and Wang, 2022). Among these, DCT-PLS method stands out for its high computational efficiency and proven success in global reconstructions, including surface currents in coastal waters (Fredj et al., 2016), Chl *a* in the Adjacent Luzon Strait (Wang et al., 2022a), and global phytoplankton functional type (Zhang et al., 2024b).

Here, we present a new consistent global dataset of POC, PON, and POC:PON at daily 4 km resolution spanning 1998–2023, addressing critical gaps in existing datasets. Daily global POC and PON concentrations were retrieved using GPR models that integrate OC-CCI ocean color products with the Global Ocean Physics Reanalysis dataset. Missing values were then reconstructed using DCT-PLS, producing gap-free global fields that outperform two commonly used reconstruction methods.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

The dataset was rigorously validated against independent in-situ observations, and the POC product was further evaluated through TCA against two widely used global POC products. Comparisons of spatiotemporal variability highlight the necessity of gap filling and demonstrate the improved capability of the dataset to consistently represent global POM dynamics across space and time. The reconstructed products provide continuous global POC and PON distributions, and particularly constitute a self-consistent POC:PON dataset, enabling systematic investigation of the biogeochemical, physiological, and ecological drivers of POC:PON variability.

2 Materials and methods

2.1 General framework

The generation of global gap-free daily POM data products from ocean color observations is challenged by the complex nonlinear relationships between optical signals and biogeochemical properties, together with extensive missing data resulting from cloud contamination and unfavorable observation conditions. In this study, surface POC and PON retrievals were formulated as a nonlinear relationship, incorporating bio-optical and physical predictors to better capture their spatial heterogeneity and environmental dependence. To address these challenges, an integrated framework was developed (Fig. 1), combining multi-source satellite observations and reanalysis data with separate retrieval and reconstruction processes.

Specifically, GPR was employed to model the nonlinear relationships between in situ measurements of POC and PON and the corresponding bio-optical and physical variables. This approach provided robust estimates. To achieve spatiotemporally continuous daily fields, the retrieved POC and PON data were reconstructed using the DCT-PLS approach, effectively filling missing values while preserving spatial structures. Using the proposed framework, a global gap-free daily 4 km (8640 × 4320 pixels) POC, PON, and POC:PON was produced for the period from 1 January 1998 to 31 December 2023.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

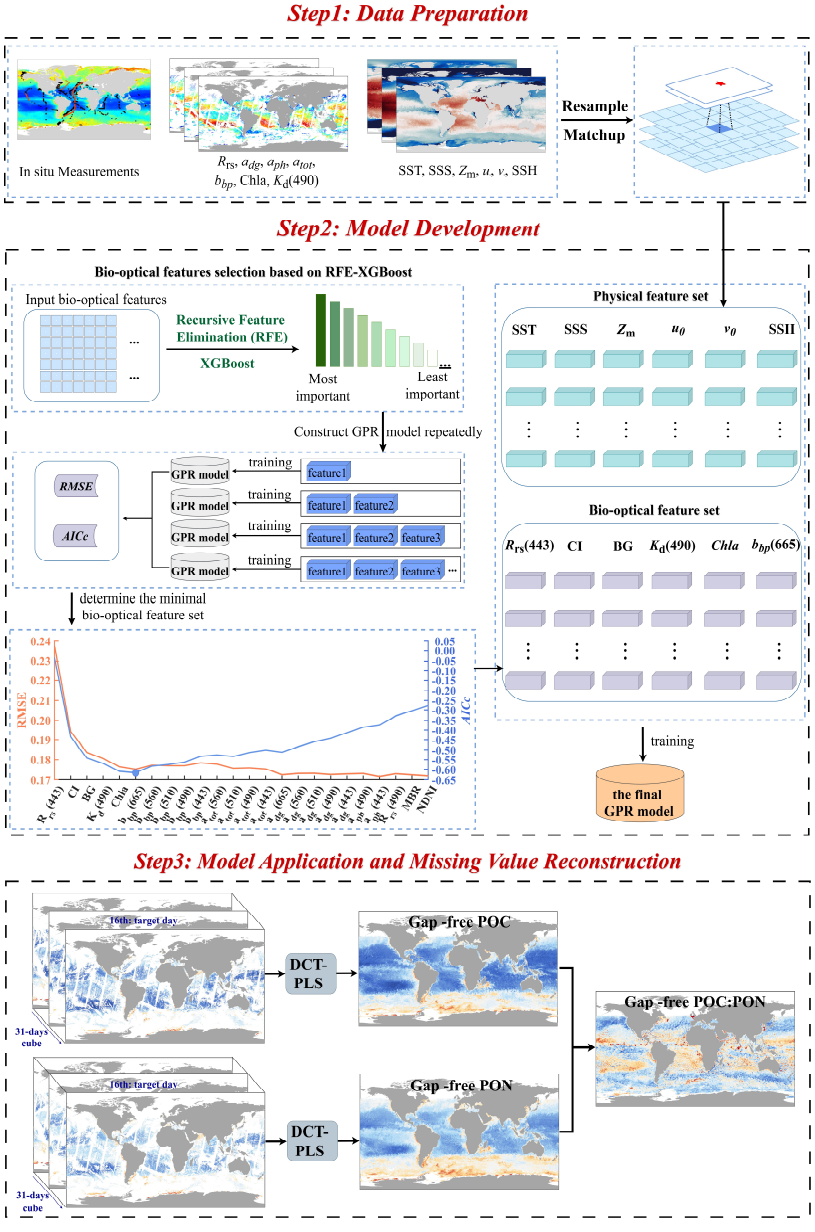


Figure 1. Workflow for the generation of the global gap-free POC, PON, and POC:PON datasets.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

120

2.2 Data sources and preprocessing

2.2.1 In situ POC and PON observations

In-situ POC and PON were collected from three publicly accessible databases, including the SeaWiFS Bio-optical Archive and Storage System (SeaBASS, https://seabass.gsfc.nasa.gov/; (Werdell and Bailey, 2005)), the Dryad Digital Repository
125 (https://datadryad.org/; Martiny et al. (2014)), and Zenodo (https://zenodo.org/; Tanioka et al. (2022)). These sources provide extensive measurements of POC and PON concentrations across the global ocean. After vertical averaging within the upper 10 m and removal of duplicate records, 11,050 surface POC and 12,390 surface PON observations were retained.

2.2.2 Satellite ocean color dataset

Ocean color products from OC-CCI Version 6.0 (available at http://www.esa-oceancolour-cci.org), provided by the
130 European Space Agency, were used as the primary source for retrieving POC and PON concentrations. This dataset provides daily 4km merged ocean color observations from SeaWiFS, MERIS, MODIS, VIIRS, Sentinel 3A/B OLCI. During the merging procedure, observations from the contributing sensors were spectrally adjusted and bias-corrected to ensure consistency with the MERIS data, followed by pixel-level uncertainty estimation. Despite multi-sensor integration and uncertainty quantification, gaps in coverage remain, requiring reconstruction to create continuous POC and PON datasets.

135 #### 2.2.3 Global ocean physics reanalysis dataset

Daily Global Ocean Physics Reanalysis data (GLORYS12V1, 2001–2022; https://doi.org/10.48670/moi-00021) were acquired from the Copernicus Marine Data Store (https://data.marine.copernicus.eu/products) to capture spatiotemporal variability in oceanic physical properties. This high-resolution dataset (1/12°, approximately 8 km) includes key physical variables such as sea surface temperature (SST) and sea surface salinity (SSS). The product is produced by a global
140 operational forecasting system that integrates multiple observational datasets, including along-track altimetry-derived sea level anomaly, satellite SST, and in-situ temperature and salinity profiles. This product provides a physically coherent and dynamically realistic reconstruction of ocean conditions.

2.2.4 Spatiotemporal matching of in-situ, satellite, and reanalysis dataset

A spatiotemporal collocation procedure was applied to match the in-situ POC and PON observations with the OC-CCI
145 dataset. For each in-situ sample, the nearest OC-CCI pixel was assigned for collocation. Satellite observations within a 3×3 pixel neighborhood and a temporal of ± 1 day were then extracted. Matchups were accepted only when at least five pixels within the window contained valid data. And the satellite observation was represented by the mean of valid pixels (Liu et al., 2021). The matched OC-CCI records were then combined with physical variables from the GLORYS12V1 reanalysis dataset,

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

after resampling all GLORYS12V1 variables to the spatial resolution of OC-CCI products. The same spatial (3 × 3 pixels) and temporal (±1-day) criteria adopted for the OC-CCI matchups were also applied to the reanalysis data.

This procedure yielded 3110 and 3118 valid matchups for POC and PON, respectively. The geographical distributions of these matchups are shown in Fig. S1 a–b in the supplement (black circles). Overall, the sampling sites are globally distributed, with the majority located in the Atlantic, Pacific and Indian Oceans, and a smaller fraction in the Southern Ocean. Statistical summaries of all matched samples are provided in Table S1 in the Supplement.

2.3 Satellite retrieval models for POC and PON concentrations

2.3.1 Retrieval model development

Four categories of parameters (Table 1) were evaluated as candidate predictors for retrieving surface POC and PON concentrations. AOPs and IOPs at 412 nm were excluded due to the high uncertainty associated with imperfect atmospheric correction (Wei et al., 2020). To identify the optimal bio-optical features for POC and PON retrievals, a recursive feature elimination approach based on the XGBoost algorithm (RFE-XGBoost) was employed. Predictors were ranked by importance, and the optimal feature subset was identified based on the minimum value of the Akaike Information Criterion with small-sample correction (*AICc*). A detailed description of this feature-selection procedure can be found in Zhang et al. (2024c).

In addition to bio-optical variables, the physical properties listed in Table 1 were incorporated into the retrieval models. These variables were not subjected to the RFE-XGBoost procedure because their influence on POM is mediated through complex, indirect dynamical processes, often resulting in weak linear correlations with POC and PON concentrations. Consequently, they would have been removed early in the feature selection process, despite their potential contributions to model accuracy and physical consistency.

Table 1. Candidate predictors employed in the retrieval models of surface POC and PON.

Dataset	Parameter type	Abbreviation	Definition	Resolution
OC-CCI dataset	Apparent optical properties	$R_{rs443\text{-}665}$	remote sensing reflectance at 443, 490, 510, 560, and 665 nm	~4 km, daily, 1 Jan 1998–31 Dec 2023
		$K_d(490)$	diffuse attenuation coefficient of downwelling irradiance at 490 nm	
		BG	blue-to-green reflectance ratio: $R_{rs}(\lambda_b)/R_{rs}(\lambda_g)$	
		CI	color index: $R_{rs}(\lambda_g)$-[$R_{rs}(\lambda_b)$+(λ_g-λ_b)/(λ_r-λ_b)]×[$R_{rs}(\lambda_r)$-$R_{rs}(\lambda_b)$]	

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

		MBR	maximum band ratio: max[$R_{rs}(\lambda_{b1})$, $R_{rs}(\lambda_{b2})$, …, $R_{rs}(\lambda_{bn})$]/$R_{rs}(\lambda_g)$	
		NDNI	normalized difference nitrogen index: [$R_{rs}(\lambda_g)$-$R_{rs}(\lambda_b)$]/[$R_{rs}(\lambda_g)$+$R_{rs}(\lambda_b)$]	
	Inherent optical properties	$a_{tot443\text{-}665}$	total absorption coefficient at 443, 490, 510, 560, and 665 nm	
		$a_{dg443\text{-}665}$	absorption coefficient of detritus and gelbstoff at 443, 490, 510, 560, and 665 nm	
		$a_{ph443\text{-}665}$	absorption coefficient of phytoplankton at 443, 490, 510, 560, and 665 nm	
		$b_{bp443\text{-}665}$	total backscattering coefficient at 443, 490, 510, 560, and 665 nm	
	Biological properties	*Chl-a*	Chlorophyll-a concentration	
GLORYS12V1 dataset	Physical properties	SST	Sea surface temperature	1/12°, daily, 1 Jan 1998–31 Dec 2023
		SSS	Sea surface salinity	
		Z_m	Mixed layer depth	
		u_0	Eastward velocity	
		v_0	Northward velocity	
		SSH	Sea surface height	

The matchup dataset was randomly split into calibration (80%) and validation (20%) subsets for model development (Table S1 in the Supplement). GPR, a nonparametric kernel-based method, was chosen for model training as it outperformed parametric models like neural networks in preliminary experiments. The model's robustness was evaluated using ten-fold cross-validation. Moreover, an independent in situ dataset (illustrated as orange circles in Fig. S1 a-b in the Supplement) from https://www.bco-dmo.org/dataset/526747 (Martiny et al., 2014), collected during the AE1206, AE1319, NH1418, and AMT24 cruises and not used in model development, was used to assess the accuracy of the retrieval models.

2.3.2 Model interpretation method

To interpret the GPR POC and PON retrieval models, SHapley Additive exPlanations (SHAP) were used to assess the contribution of individual features on the model outputs, including both the strength and direction of their effects. For each observation, SHAP analysis decomposes the model output into contributions from each feature, with a SHAP value

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

indicating the effect of each predictor on the corresponding prediction. Here, $\emptyset_{i,j}$ denotes the SHAP value of the j-th feature for the i-th sample, and the corresponding prediction is given by:

$$\hat{y}_i = \emptyset_0 + \sum_{j=1}^{k} \emptyset_{i,j} , \quad (1)$$

where $\emptyset_0$ is the model baseline output, typically defined as the average of the target variable over all samples; and k represents the total number of input features in the model. Positive SHAP values indicate that a feature increases the predicted value, whereas negative SHAP values suggest a reduction in the prediction.

SHAP analysis provides both global and local interpretability of the GPR models. Global interpretability evaluates the overall influence of each input feature on model predictions, while local interpretability reveals the contribution of individual features to specific predictions, thereby elucidating the model decision-making process and the influence of each feature on individual samples.

2.4 Reconstruction of missing values in satellite-retrieved POC and PON concentrations

2.4.1 Reconstruction strategy

Daily global POC and PON concentrations were initially retrieved using GPR models. However, missing values required reconstruction to produce gap-free fields. Several methods, such as DINEOF (Beckers and Rixen, 2003), DINCAE (Barth et al., 2020), and OI (Reynolds and Smith, 1994), have been proposed for this task, but they can be computationally demanding for large-scale, long-term datasets. In this study, a robust, iteratively weighted version of the DCT-PLS algorithm was adopted. Owing to its computational efficiency, modest memory demands, and reliable reconstruction performance, the method is well suited for large spatiotemporal datasets. DCT-PLS efficiently smooths one- and multi-dimensional data by down-weighting potential outliers while giving higher weight to reliable observations (Garcia, 2010).

The reconstruction process was designed as follows: (1) Spatiotemporal cube construction. For each target day t, a three-dimensional spatiotemporal data cube was constructed by stacking daily POC and PON fields within a ±15-day temporal window centered on t, resulting in a 31-day cube (4320 × 8640 × 31). This window length was chosen to capture short-term temporal continuity while maintaining computational efficiency. The reconstruction window for the first fifteen days of January 1998 and the last fifteen days of December 2023 adopted the most recent 31-day cube that contains the target day. (2) Data reconstruction via DCT-PLS. The DCT-PLS method was applied to each cube to estimate missing values, utilizing both spatial and temporal correlations. The process was iterated 100 times to ensure convergence. (3) Target Day Extraction. After reconstruction, the central temporal slice corresponding to day t was extracted as the final gap-free POC or PON data. This procedure was repeated for each day, ultimately generating a spatially complete, temporally continuous daily global POC and PON datasets spanning 1998–2023.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

2.4.2 Evaluation of reconstruction performance

The reconstruction performance of DCT-PLS was evaluated using two approaches: (1) Gap simulation experiment. The accuracies of DCT-PLS, DINEOF, and DINCAE were compared in nine randomly selected areas (indicated by the red diamond in Fig. S2 in the Supplement) across representative oceanic regions defined by the Regional Carbon Cycle Assessment and Processes framework (https://reccap2-ocean.github.io/regions/; Canadell et al. (2011)). For each area, 10% of valid pixels for the target day were randomly removed and treated as missing values, which were then reconstructed and compared with the original satellite-retrieved data to evaluate reconstruction accuracy. (2) Independent validation using in situ data. An independent in situ dataset (https://www.bco-dmo.org/dataset/526747), previously used to evaluate the GPR retrieval model in Sect.2.3, was also employed to assess the performance of DCT-PLS.

TCA was further applied to evaluate the performances of the GGFD POC product against two widely used global POC products, MODIS-Aqua and MULTIOBS (Table 2). The application of TCA relies on three key assumptions (Kim et al., 2023): each dataset is linearly related to the underlying true signal, the associated errors are orthogonal to the signal, and error terms from different datasets are mutually independent. Adherence to these principles ensures that TCA delivers a reliable and unbiased evaluation of errors and product quality.

Monthly means of all POC datasets were calculated and interpolated onto a common 1° × 1° grid. These fields were then concatenated into a continuous time series for TCA analysis, which quantified relative errors and correlations of the three datasets with respect to the unknown true POC. The fractional mean-squared error (*fMSE*) and the squared correlation coefficient were used to express the error statistics derived from TCA. The *fMSE* was calculated as:

$$fMSE_i = \frac{\sigma_{\varepsilon_i}^2}{\sigma_i^2} = \frac{\sigma_{\varepsilon_i}^2}{\beta_i^2 \sigma_\Theta^2 + \sigma_{\varepsilon_i}^2} = \frac{1}{1+\mathrm{SNR}_i}, \quad (2)$$

where i represents one of the three datasets; $\sigma_{\varepsilon_i}^2$ denotes the error variance estimated by TCA for dataset; β_i is the scaling coefficient; σ_i^2 represents the variance of dataset i; σ_Θ^2 corresponds to the variance of the underlying true POC signal; and SNR denotes the signal-to-noise ratio (Zhang et al., 2024b; Kim et al., 2023). The squared correlation coefficient (R_i^2) can be expressed as:

$$R_i^2 = \frac{\beta_i^2 \sigma_\Theta^2}{\beta_i^2 \sigma_\Theta^2 + \sigma_{\varepsilon_i}^2} = \frac{\mathrm{SNR}_i}{1+\mathrm{SNR}_i}, \quad (3)$$

It can be observed that the sum of *fMSE* and R_i^2 equals 1.

Table 2. Summary of publicly released POC data products used for comparison in this study.

Product	Method	Spatial resolution	Temporal resolution	Reference
MODIS-Aqua	blue-to-green reflectance ratio algorithm	~4 km	daily	(Stramski et al., 2008)

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

MULTI OBS	Transfer from a dedicated transfer function of b_{bp} to POC	0.25°×0.25°	weekly	(Sauzède et al., 2016)

2.5 Spatiotemporal trend analysis

To detect long-term trends in POC, PON and POC:PON, the climatological annual signal at each grid cell was first subtracted according to the method described by Vantrepotte and Mélin (2009). This annual cycle was computed from the gap-free daily data products as the mean for each calendar month over 1998–2023. Linear regressions were then fitted to the deseasonalized time series at each grid cell, with the slopes representing trends expressed as percent per year (% year^{-1}). Statistical significance was assessed using p-value, and regions with non-significant trends ($p > 0.05$) were masked in white (Pauthenet et al., 2024).

To characterize the global evolution of oceanic POM, spatially weighted monthly medians of POC, PON, and POC:PON were calculated from January 1998 to December 2023, with weights proportional to the cosine of latitude. Medians were preferred over means because they are less sensitive to outliers and skewed distributions (Wasserman, 2013).

2.6 Uncertainty propagation analysis of POC:PON

The impact of uncertainties in gap-free POC and PON datasets on the POC:PON ratio was analyzed using uncertainty propagation theory (Lee et al., 2010). The uncertainty of POC:PON (ΔCNR) was calculated considering both retrieval and reconstruction processes. For each process, ΔCNR is expressed as a function of the uncertainties in POC (ΔPOC) and PON (ΔPON):

$$\Delta\mathrm{CNR} = \sqrt{\frac{\Delta\mathrm{POC}^2}{\mathrm{PON}^2} + \frac{\mathrm{POC}^2\Delta\mathrm{PON}^2}{\mathrm{PON}^4}}, \quad (4)$$

3. Results

3.1 POC and PON retrieval models

3.1.1 Performance of POC and PON retrieval models

Six bio-optical features were selected from 24 candidates for satellite retrieval of POC and PON concentrations, guided by the lowest *AICc* values and relatively low RMSE (Fig. S3 in the Supplement). The selected features for both POC and PON were $R_{rs}(443)$, CI, BG, $K_d(490)$, Chl-*a*, and $b_{bp}(665)$. Model RMSE decreased as additional features were added until reaching a plateau, indicating that the minimum *AICc* criterion effectively reduced model complexity without increasing prediction error.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

The GPR POC and PON models exhibited high performance for both calibration and validation datasets (Fig. 2). For the calibration of both POC and PON, the models achieved comparable performance, with R^2 values of 0.96 and RMSE values of 1.23 mg m^{-3}. The MAPE were 3.32% for POC and 6.26% for PON. The regression exhibited slopes of approximately 0.95, indicating the models slightly overpredicted lower values while tending to underestimate higher concentrations. The cross-validation analysis further demonstrated stable model performance. For POC, prediction accuracy was characterized by an R^2 of 0.88, an RMSE of 1.42 mg m^{-3} and a MAPE of 5.76%. Corresponding values for PON were 0.87, 1.45 mg m^{-3}, and 11.32%, respectively. Slopes of 0.94 obtained during cross-validation further supported the consistency between observations and model estimates.

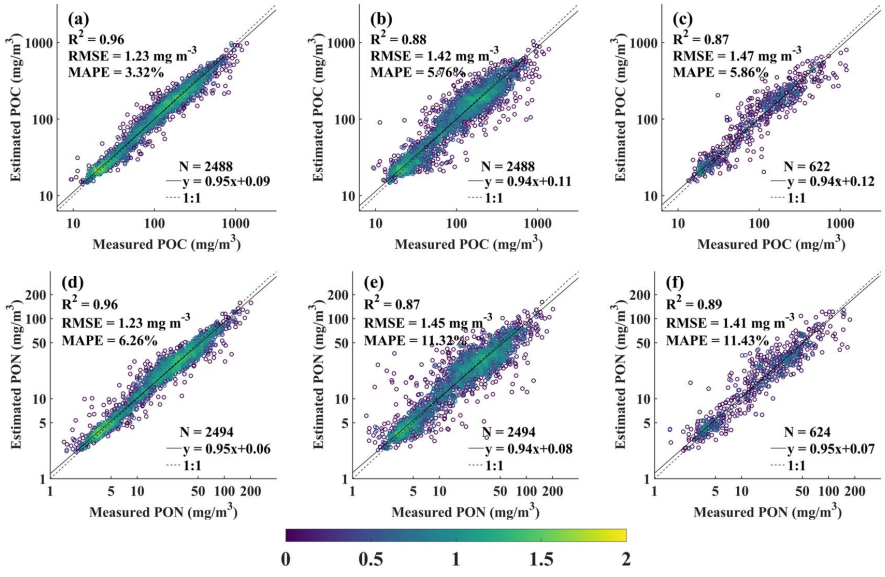


Figure 2. Scatterplots of GPR-retrieved versus in-situ observations of POC (a–c) and PON (d–f): calibration (a, d), 10-fold cross-validation (b, e), and validation (c, f).

Results from the validation confirmed the strong generalization capability of the model. For POC, the model achieved an R^2 of 0.87, an RMSE of 1.47 mg m^{-3}, a MAPE of 5.86%, and a regression slope of 0.94. For PON, the corresponding values were 0.89, 1.41 mg m^{-3}, 11.43%, and 0.95, respectively. Overall, the models performed comparably for POC and PON, with slightly higher errors observed for PON. These results demonstrate that the GPR-based retrieval framework provides robust and accurate estimates of global oceanic POC and PON.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

3.1.2 Interpretability of POC and PON retrieval models

Figure 3 shows the influence of individual features on the outputs of the GPR-based POC and PON models, illustrating both local and global interpretability. For the prediction of log(POC), the most influential predictors in descending order were log(Chla), BG, Z_m, SST, log(b_{bp}(665)), SSS, log(R_{rs}(443)), and log(K_d(490)). In the GPR POC model (Fig. 3a and b), log(Chla) exhibited the widest SHAP value distribution, indicating its dominant influence on model outputs. Consistent with the strong positive correlation between log(POC) and log(Chla) (R = 0.90; Table 3), high Chla values were associated with positive SHAP values, whereas low values contribute negatively, suggesting that increasing Chla enhanced predicted POC. This is physically reasonable, as elevated Chla reflected increased phytoplankton biomass, leading to higher concentrations of both POC and PON. BG was the second most influential predictor and exhibited an opposite SHAP pattern, with high BG contributing negatively and low BG positively, consistent with its strong negative correlation with log(POC) (R = −0.90; Table 3).

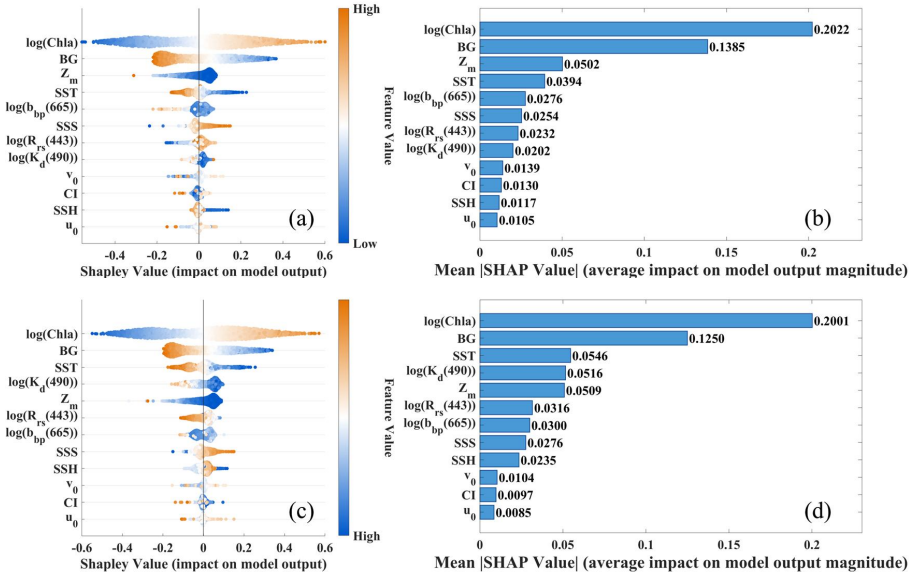


Figure 3. (a) Local and (b) global interpretability of the GPR POC model, and (c) local and (d) global interpretability of the GPR PON model.

The contribution of Z_m ranked third and showed a more complex SHAP distribution, reflecting its relatively weak negative correlation with log(POC) (R = −0.48; Table 3). Smaller Z_m values tended to enhance model outputs, probably because a shallower mixed layer promoted nutrient retention within the euphotic zone and facilitated phytoplankton growth. In contrast, larger Z_m values generally suppressed POC estimates. SST exhibited a contribution pattern similar to BG, with higher values

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Open Access Earth System Science Data Discussions

generally exerting a negative influence on POC predictions. This is consistent with the fact that elevated SST enhanced water column stratification, suppressed nutrient supply, and limited POM production.

log(b_{bp}(665)) and log(K_d(490)) exhibited similar effects, with lower values contributing positively and higher values tending to suppress model outputs. This behavior could reflect reduced sensitivity of remote-sensing retrievals to biologically derived POC signals under conditions of enhanced particle scattering and light attenuation. Notably, the SHAP contribution patterns of these variables did not fully align with their significant positive correlations with POC. This discrepancy highlights the importance of nonlinear interactions and combined modulation effects, which cannot be captured by simple linear relationships. In contrast, log(R_{rs}(443)) exhibited a distinctly non-monotonic contribution, reflecting the combined effects of phytoplankton absorption, non-algal particles, and CDOM. Other physical variables showed more dispersed and heterogeneous contributions. For example, the SHAP values of SSS spanned both positive and negative ranges, suggesting that its effects on POC are mediated through complex and interacting oceanic processes.

Similarly, for log(PON), the dominant contributors included log(Chla), BG, SST, log(K_d(490)), Z_m, log(R_{rs}(443)), log(b_{bp}(665)), and SSS. In the GPR-PON model (Fig. 3c and d), the relative importance of individual features differed moderately from that in the GPR-POC model, while the directions of their contributions remained largely consistent. The primary discrepancy occurred for log(R_{rs}(443)), where higher values were generally associated with negative SHAP values, whereas lower values contributed positively.

Table 3. Correlation coefficients between the predictors and the POC PON concentrations.

Predictors	log(Chla)	BG	Z_m	SST	log(b_{bp}(665))	log(K_d(490))	log(R_{rs}(443))	SSS	V_0	SSH	CI	U_0
log(POC)	0.90	-0.90	-0.48	-0.67	0.81	0.87	-0.83	-0.70	0.10	-0.56	0.88	0.04
Predictors	log(Chla)	BG	log(K_d(490))	SST	Z_m	log(R_{rs}(443))	log(b_{bp}(665))	SSS	SSH	V_0	CI	U_0
log(PON)	0.90	-0.89	0.87	-0.68	-0.47	-0.83	0.80	-0.69	-0.55	0.09	0.88	0.03

3.2 Evaluation of DCT-PLS for POC and PON reconstruction

3.2.1 Comparison of DCT-PLS, DINEOF, and DINCAE

DCT-PLS demonstrated consistently superior performance in reconstructing both POC and PON concentrations, with high accuracy maintained across different seasons. To illustrate this performance in a representative region, the reconstruction results of the Equatorial Atlantic (marked in Fig. S2 in the Supplement) is presented as a case study. For POC, R^2 exceeded 0.96 and RMSE remained close to 1.03 mg m^{-3} for both 1st January and 1st July 2020 (Fig. 4a and b). For PON, R^2 ranged from 0.95 to 0.97, with similarly low RMSE of 1.05 mg m^{-3} and 1.03 mg m^{-3}, respectively (Fig. 4c and d). The MAPE for PON was higher than that for POC, with values of 1.57% and 1.03% for PON, and 0.54% and 0.46% for POC, corresponding to 1st January and 1st July 2020, respectively. The reconstruction accuracy for PON was therefore slightly lower than that for POC. But overall, both variables were reconstructed with high quality, highlighting the robustness of the DCT-PLS approach.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

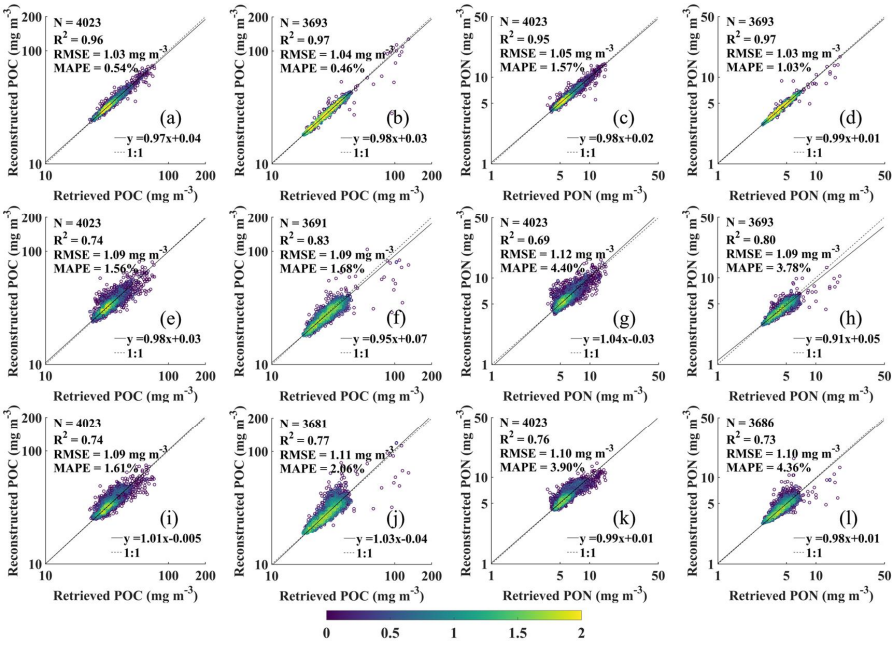


Figure 4. Scatterplots of reconstructed versus GPR-retrieved POC and PON for 1st January and 1st July 2020. Panels (a)-(d) correspond to DCT-PLS (POC: a,b; PON: c,d), panels (e)-(h) to DINEOF (POC: e,f; PON: g,h), and panels (i)-(l) to DINCAE (POC: i,j; PON: k,l). All concentrations are shown on a log10 scale, and the color bar indicates point density.

In comparison, DINEOF showed noticeably reduced performance. For POC, R^2 decreased to 0.74 and 0.83, accompanied by MAPE larger than 1.50% (Fig. 4e and f). For PON, the degradation was more pronounced, with R^2 dropping to as low as 0.69 on 1st January and MAPE both exceeding 3.70% on 1st January and 1st July (Fig. 4g and h). These results indicate that DINEOF struggled to accurately capture the variability of PON compared to DCT-PLS.

DINCAE exhibits a different behavior. While its performance for POC is comparable to that of DINEOF on 1st January with the same R^2 and RMSE, and a slightly higher MAPE of 1.61% (Fig. 4i). And for 1st July, the accuracy of DINCAE was lower than that of DINEOF, with decreased R^2 and increased RMSE and MAPE (Fig. 4j). Relative to DINEOF, DINCAE improved PON reconstruction performance on 1 January, increasing the R^2 to 0.76 and reducing the RMSE and MAPE to 1.10 mg m^{-3} and 3.90%, respectively (Fig. 4k and l). This suggests that DINCAE could better capture certain nonlinear features in PON variability than DINEOF, although its overall performance remained inferior to DCT-PLS.

A comprehensive comparison across all nine regions further confirms the robustness of DCT-PLS (Table S2 in the Supplement). The DCT-PLS method consistently outperformed both DINEOF and DINCAE in terms of reconstruction accuracy. DINCAE achieved performance comparable to that of DINEOF, but its substantially higher computational cost greatly constrains its applicability for large-scale ocean data reconstruction. In contrast, DCT-PLS exhibited clear advantages

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

in computational efficiency, making it a more practical and scalable approach for large-scale applications, and was therefore adopted in this study.

3.2.2 Performance of DCT-PLS reconstruction across the global ocean

Using the DCT-PLS method, the majority of missing oceanic POC and PON observations were recovered by exploiting the spatiotemporal continuity and intrinsic correlations within a 31-day dataset centered on the target date. Fig. S4 in the Supplement illustrates the daily number of valid pixels for both POC and PON before and after reconstruction. The original satellite-retrieved products exhibited substantial data gaps, with valid pixels accounting for only approximately 24.21% of the final reconstructed dataset on average. This result indicates that the DCT-PLS method increases data availability by nearly fourfold, substantially reducing the inherent sparsity in satellite-derived POC and PON concentrations.

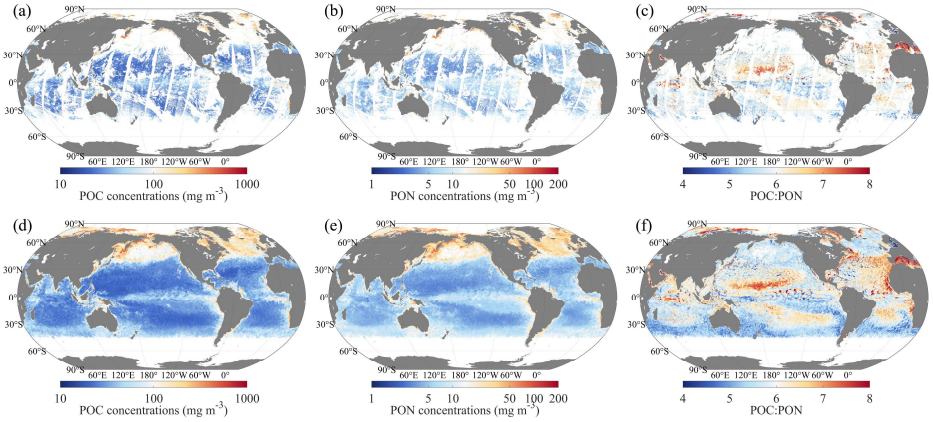


Figure 5. Global oceanic distributions of satellite-retrieved and reconstructed POC concentration (a, d), PON concentration (b, e), and POC:PON ratios (c, f) products on 1 July 2020.

To further evaluate reconstruction performance, the spatial completeness and continuity of the POC and PON fields were examined. As shown in Fig. 5, it presents global distributions of satellite-retrieved and reconstructed POC, PON and POC:PON on 1 July 2020, the reconstructed maps more clearly revealed large-scale spatial patterns compared with the original retrieved maps. Specifically, the number of valid pixels increased from 5,061,773 in the retrieved dataset to 15,782,282 after reconstruction. This substantial increase demonstrates that the DCT-PLS method effectively mitigated spatial discontinuities caused by cloud contamination and observational gaps, while preserving the underlying large-scale distribution features of POM.

Nevertheless, a small portion of POM data remained unreconstructed in regions persistently lacking satellite observations, such as areas affected by polar night, high solar zenith angles, or frequent cloud cover (e.g., red-boxed area in Fig. 6a). This limitation reflects an intrinsic constraint of gap filling when observational data are absent from extended periods. In practice,

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

reconstruction of a given pixel within the 31-day window was often not possible if fewer than five valid observations were available (Fig. S5 in the Supplement), also depending on the availability of valid data in neighboring pixels.

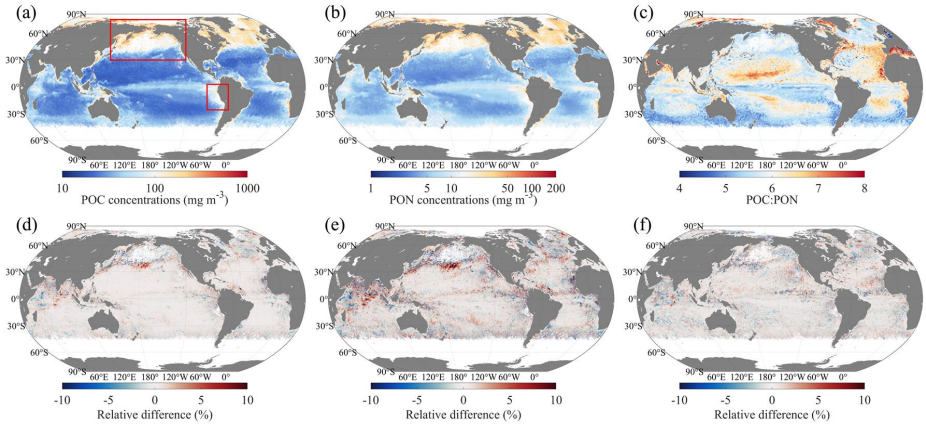


Figure 6. Monthly mean gap-free maps of POC, PON, and POC:PON (a-c), and corresponding relative differences between the reconstructed and satellite-retrieved values for July 2020 (d-f).

Figure 6 shows the monthly mean gap-free maps of POC, PON, and POC:PON (Fig. 6a–c), together with maps of the unbiased relative difference (URD, URD = $(y_1-y_2)/(y_1+y_2) \times 200\%$) between the reconstructed and satellite-retrieved monthly averages for July 2020 (Fig. 6d–f). The gap-free POC exhibited strong consistency with the satellite retrievals, with URD values ranging from -41.58% to 58.55% and a mean of 0.10%±1.55%. In contrast, the URD for PON ranged from -70.45% to 124.97%, with a mean of -0.24%±2.82%, indicating substantially larger variability. The derived POC:PON ratio showed the smallest variability, ranging from -41.60% to 40.78%, with a mean of -0.003%±1.89%.

Spatially, the reconstructed and satellite-retrieved results exhibited broadly consistent global patterns, with URD values close to zero across oceanic regions. Pixel-wise comparisons between the monthly mean fields showed strong agreement, suggesting that the DCT-PLS approach effectively preserves the spatial patterns and statistical characteristics of POC and PON. Noticeable positive and negative URD patches were observed in the northern North Pacific, North Atlantic, and Northern Indian Oceans for all the three variables. These discrepancies were more widespread and pronounced in the PON dataset, further indicating its relatively higher reconstruction uncertainty.

3.3 Independent evaluation of satellite retrieval and missing value reconstruction

To objectively evaluate the accuracy of both the satellite-retrieved POM results and the reconstructed POM products, an independent in situ dataset, which was not involved in model development, was used. These samples were mainly obtained from oligotrophic regions of the Atlantic and Pacific Oceans, with a few from coastal waters of the Atlantic. The in situ POC

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

concentrations ranged from 13.09 to 220.18 mg m^{-3}, and PON concentrations ranged from 1.54 to 33.01 mg m^{-3} (Fig. S1 c–d in the Supplement).

Scatterplots comparing satellite-retrieved and reconstructed POC, PON, and POC:PON against in situ measurements are shown in Fig. 7, based on 60 satellite–in situ matchups. Overall, both POC and PON exhibited strong agreement with in situ observations, with R^2 of 0.73 and 0.81, RMSE of 1.56 mg m^{-3} and 1.51 mg m^{-3}, and MAPE of 6.02% and 15.59%, respectively. These results suggest that satellite-retrieved POC was estimated more accurately than PON. In contrast, the POC:PON ratio exhibited lower predictive performance, with an R^2 of 0.43, RMSE of 1.07 mg m^{-3}, and MAPE of 11.69%. Regression slopes for retrieved POC, PON, and POC:PON were 0.65, 0.70, and 0.67, respectively, indicating overestimation at low and underestimation at high values.

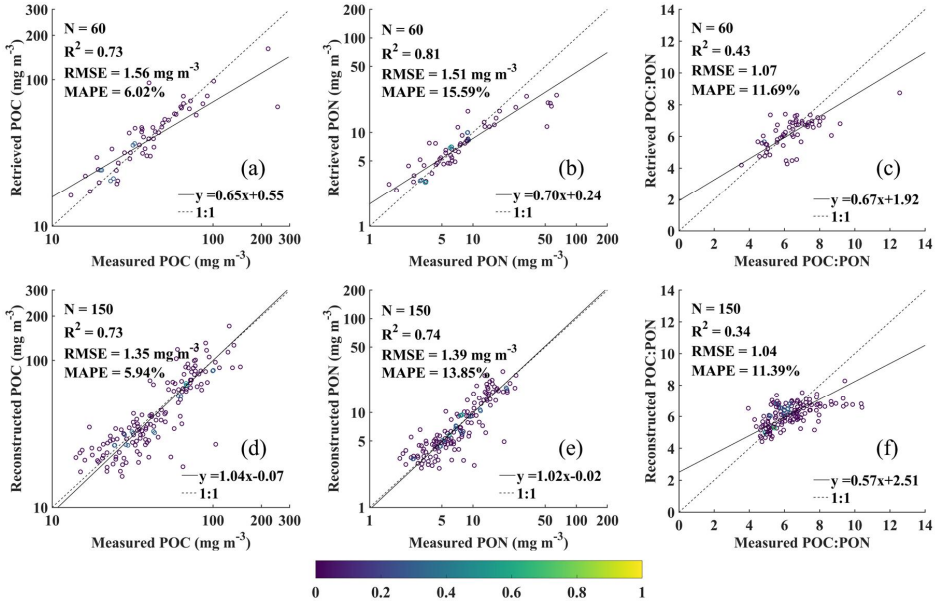


Figure 7. Scatterplots of satellite-retrieved versus in situ measured values for POC (a), PON (b), and POC:PON (c), and scatterplots of reconstructed versus in situ values for POC (d), PON (e), and POC:PON (f) based on an independent validation dataset. The color bar represents point density.

For the reconstructed data, a total of 210 matchups with in situ observations were available. To ensure an independent evaluation, the 60 matchups used for satellite retrieval validation were excluded, leaving 150 matchups for assessing reconstruction performance. Overall, reconstruction performance was slighter better for POC than for PON, with R^2 of 0.73 and 0.74, RMSE of 1.35 mg m^{-3} and 1.39 mg m^{-3}, and MAPE of 5.94% and 13.85%, respectively. The reconstructed POC:PON ratio exhibited lower accuracy, with an R^2 of 0.34, RMSE of 1.04, and MAPE of 11.39%. The reconstructed POC

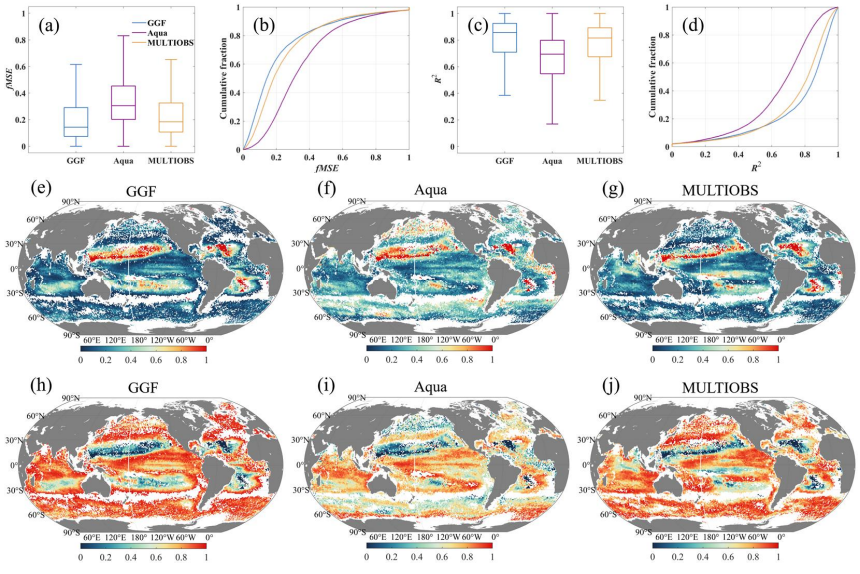
https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

and PON data showed improved performance compared to retrieval results, with regression slopes of 1.04 and 1.02, respectively. However, the POC:PON ratio still exhibited a tendency to be overestimated at low concentrations and underestimated at high concentrations, with a regression slope of 0.57.

Overall, both satellite-retrieved and reconstructed POC concentrations showed relatively good accuracy, followed by PON concentrations. The performance for POC:PON was lowest, integrating the compounded uncertainties from both POC and PON measurements.

3.4 Comparison of GGFD POC products with MODIS-Aqua and MULTIOBS

Because of the coarse temporal resolution of the MULTIOBS product, TCA analysis was performed on the monthly POC from GGFD, MODIS-Aqua, and MULTIOBS datasets to demonstrate the superior performance of the GGFD dataset developed in this study. The statistical results of R^2 and *fMSE* derived from the TCA analysis were showed in Fig. 8a-d. Overall, monthly GGFD POC outperformed the other two products, exhibiting the highest median R^2 (0.86) and the lowest median *fMSE* (0.14). The *fMSE* cumulative distribution function (CDF) of GGFD POC was shifted toward lower values with a gentler slope, while its R^2 CDF was shifted toward higher values and appeared steeper compared with the other two datasets. These features indicate that GGFD POC achieved lower errors and higher consistency across a larger fraction of grid points, reflecting improved spatial robustness. MULTIOBS exhibited intermediate performance, with a median *fMSE* of 0.18 and an R^2 of 0.82, whereas MODIS-Aqua POC showed the largest global errors, with a median *fMSE* of 0.31 and an R^2 of 0.69.



https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Figure 8. TCA results for GGFD, MODIS-Aqua, and MULTIOBS monthly POC products. (a)-(b) show global oceanic statistics of fMSE, and (c)-(d) show those of R^2; (e)-(g) illustrate the spatial distribution of fMSE, and (h)-(j) present the spatial distribution of R^2 for GGFD, MODIS-Aqua, and MULTIOBS, respectively.

The spatial distributions of R^2 and *fMSE* for the three POC products are illustrated in Fig. 8e-j. GGFD POC demonstrated consistently superior performance across most regions of the global ocean, although relatively elevated *fMSE* values were observed in subtropical gyre zones (Fig. 8e). Most *fMSE* values were below 0.5, indicating that the true POC signal dominates over the estimation noise (Zhang et al., 2024b), and that GGFD POC is therefore a precise product. MULTIOBS POC displayed a spatial pattern similar to that of GGFD, also showing elevated *fMSE* in subtropical gyres (Fig. 8g). In contrast, MODIS-Aqua POC showed lower *fMSE* in subtropical gyres, particularly in the southern Pacific, but exhibited substantially higher errors in mid- to high-latitude regions than the other two datasets (Fig. 8f). The spatial pattern of R^2 for GGFD and MULTIOBS were similar, with GGFD consistently achieving higher values (Fig. 8h and j). Across the global ocean, R^2 values for MODIS-Aqua were significantly lower than those of the other two datasets, especially in the mid- to high-latitude regions (Fig. 8i). In general, the evaluation indicated that the GGFD POC outperformed the alternatives, demonstrating a clear advantage.

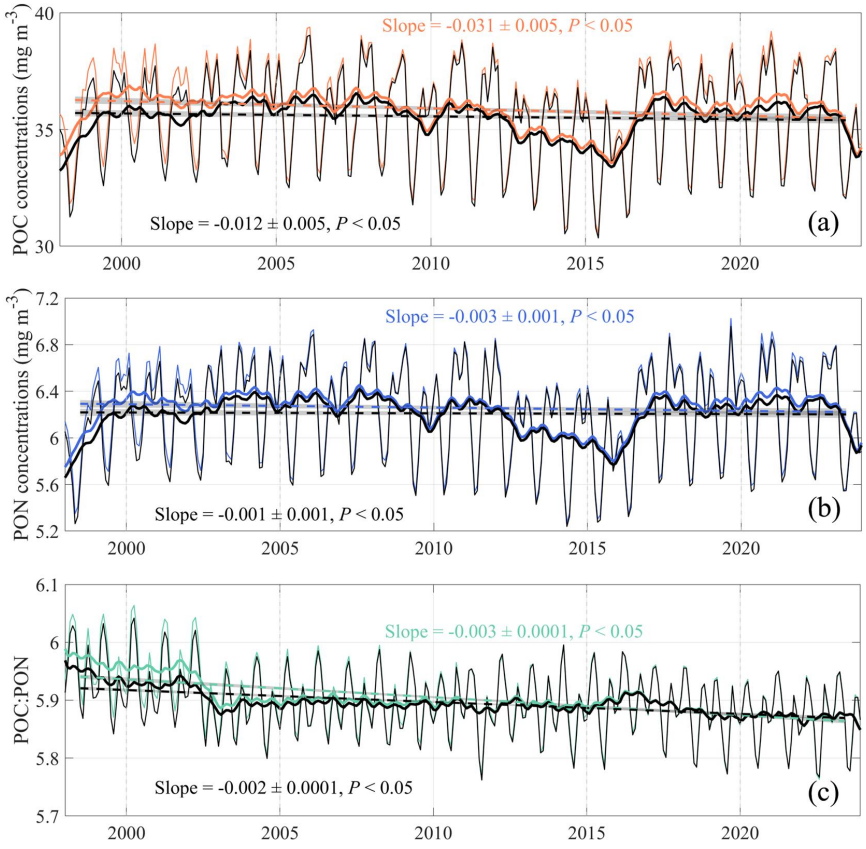
The technical framework used to generate the PON product was essentially identical to that of POC, and previous evaluations have demonstrated comparable retrieval and reconstruction accuracies for both variables. Although no global oceanic PON products are currently available for direct comparison, the TCA results for POC provide a useful reference for assessing the reliability of the PON and POC:PON products.

4 Discussion

4.1 Spatiotemporal comparison of POC, PON, and POC:PON between satellite retrievals and reconstructions

To examine the temporal trends of POC, PON, and POC:PON from satellite retrievals and reconstructions, spatially weighted medians from both datasets are presented in Fig. 9. Both POC and PON exhibited clear seasonal cycles, with annual maxima typically occurring in October and minima in June. In contrast, the POC:PON ratio followed an earlier seasonal cycle, with maxima in April and minima in August, approximately two months earlier than those of POC and PON. Pronounced interannual variations were also evident, underscoring the importance of comparing long-term trends between retrieved and reconstructed data.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026



445 **Figure 9. Spatially weighted monthly median of POC, PON, POC:PON from January 1998 to December 2023. The thin lines show the monthly median values, while thick lines represent the locally smoothed medians using a 13-month sliding window. The dotted lines represent the linear fit, the shading indicates the 95% confidence interval.**

Although the seasonal patterns were broadly consistent across the two datasets, differences in magnitudes and long-term trends were apparent. For POC, the reconstructed monthly median values and their corresponding locally smoothed medians
450 were consistently higher than those from the satellite retrievals, particularly during 1998–2003. Consequently, the overall decreasing trend of the locally smoothed median from 1998 to 2023 was more pronounced in the reconstructed data (Fig. 9a, regression slope = −0.031±0.005, $P < 0.05$) than in the satellite retrievals (Fig. 9a, regression slope = −0.012±0.005, $P < 0.05$). PON revealed similar patterns, with slopes of −0.003±0.001 (Fig. 9b, $P < 0.05$) and −0.001±0.001 (Fig. 9b, $P < 0.05$) for the reconstructed and retrieved data, respectively. For the POC:PON ratio, reconstructed values were also significantly

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

higher than that of satellite retrievals, leading to a steeper long-term decrease (Fig. 9c, regression slope = −0.003±0.0001, $P < 0.05$) compared with the satellite retrievals (Fig. 9c, regression slope = −0.002±0.0001, $P < 0.05$).

The consistently higher reconstructed values and more pronounced interannual variability may be attributed to gaps or cloud contamination in the satellite data at high latitudes or during specific seasons, which can leave many high values unrecorded. By filling these missing values, the reconstructions not only increased monthly averages but also provided a more accurate depiction of long-term interannual trends. The enhanced temporal continuity and completeness of the reconstructed datasets enable more precise and continuous representations of global POC and PON distributions and dynamics, thereby facilitating investigations into the physiological and ecological mechanisms underlying POC:PON variability.

Figure 10 further shows the long-term trends of POC, PON, and POC:PON for each grid cell across the global ocean from 1998 to 2023 derived from both satellite retrievals and reconstructed data. For the reconstructions, all three variables exhibited significant declining trends, with most regions in the Pacific and Atlantic Oceans showing negative slopes (blue areas in Fig. 10a-c). In the central subtropical gyres of the North Pacific, North Atlantic, and Indian Ocean, significant positive trends (red areas in Fig. 10a and b) appeared for both POC and PON concentrations. Large white regions were also observed for both POC and PON, suggesting the absence of any statistically significant trends. In contrast, POC:PON was characterized by a scattered but widespread declining trend, with a few areas exhibiting a positive trend (Fig. 10c). The satellite retrieval revealed a similar pattern, with more white areas representing insignificant trends, which can be attributed to fewer available observations compared to the reconstructions. This to some extent highlights the advantages of reconstructed data in studying the spatiotemporal variations of POM, contributing to a more comprehensive and complete understanding.

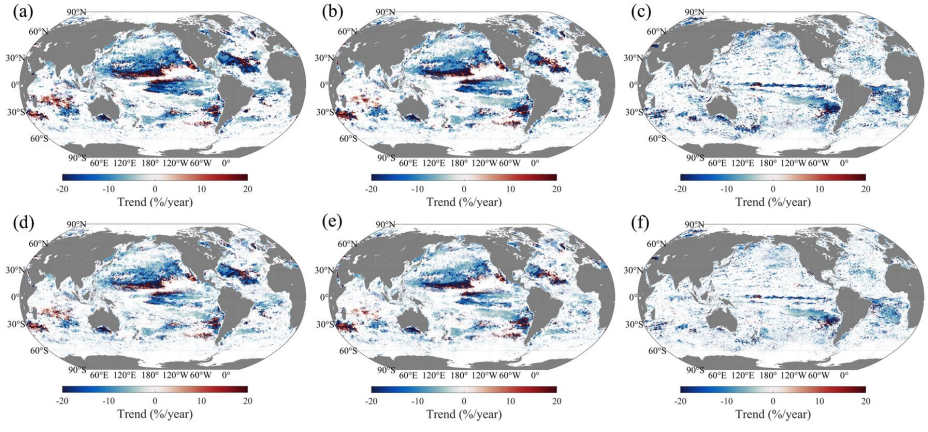


Figure 10. The linear trends of the deseasonalized POC, PON, and POC:PON for each grid cell. Trends were calculated over the period from January 1998 to December 2023. White areas indicate non-significant trends (p-value >0.05).

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

4.2 Uncertainty of POC:PON estimates and future perspectives

The final gap-free POC:PON product was derived from the gap-free POC and PON products, which were first retrieved using GPR models and subsequently reconstructed using DCT-PLS. Consequently, uncertainties arising from both the retrieval and reconstruction processes influence the accuracy of the final POC:PON ratio. To quantify these impacts, error propagation analysis was conducted using the same independent validation datasets as in Sect.3.3 for retrievals and reconstructions. It shows that uncertainties in 55/60 retrieved POC and 53/60 retrieved PON cases were less than 10% (Fig. 11a and b), and uncertainties in 145/150 reconstructed POC and 142/150 reconstructed PON cases were below 10% (Fig. 11d and e). For the POC:PON ratio, 33/60 retrieval cases were dominated by PON uncertainty, and 79/150 reconstruction cases were similarly determined by PON (Fig. 11c and f).

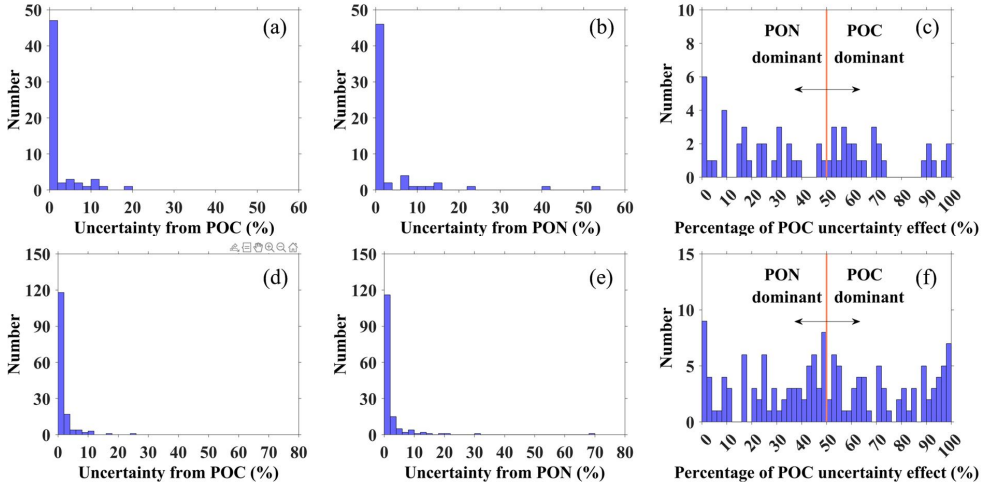


Figure 11. Uncertainty in the POC:PON derived from retrieved (a) POC and (b) PON, and reconstructed (d) POC and (e) PON, assuming zero uncertainty in other variables; and the proportion of POC uncertainty on the uncertainty of (c) retrieved and (f) reconstructed POC:PON.

Overall, the error propagation analysis indicated that the impacts of POC and PON uncertainties on the POC:PON ratio were broadly comparable, although PON uncertainty exerted a larger influence. Accuracy assessments of both satellite retrieval and reconstruction algorithms consistently showed that POC is estimated more accurately than PON. This could be due to higher natural variability of PON concentrations, which presents additional challenges for accurate satellite retrieval and reconstruction.

Future improvements could focus on several aspects. One priority is the refinement of satellite retrieval models to achieve more accurate estimates of POC and PON concentrations. In particular, expanding in-situ observations in regions that are currently sparsely sampled would help improve model generalization and applicability at the global scale. Moreover, the current GPR models are based primarily on statistical relationships and do not explicitly represent biological processes,

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

thereby limiting their ability to elucidate the mechanisms governing variations in POC, PON, and POC:PON. Integrating ecological models with remote sensing data could enhance both accuracy and interpretability. Second, the physical reanalysis data at 1/12° resolution were resampled to 4 km to match the OC-CCI product, ensuring dataset consistency. This upsampling process may affect the statistical characteristics and introduce errors. Future work could incorporate higher-resolution data and minimize information loss during processing to further improve the quality of the gap-free datasets.

5 Conclusion

This study presents the first consistent global gap-free daily 4 km dataset of POC, PON and their ratio spanning 1998–2023, providing a robust data foundation for investigating POM dynamics in relation to the carbon cycle and biogeochemical stoichiometry. GPR models were developed to retrieve POC and PON by integrating selected bio-optical and physical variables, while missing values were reconstructed using the DCT-PLS algorithm to generate gap-free global fields. Validation against independent in-situ observations confirms high accuracy for retrieved and reconstructed POC and PON, while the POC:PON ratio shows higher but still acceptable uncertainties. TCA analysis between GGFD, MODIS-Aqua and MULTIOBS POC products further demonstrates the superior performance of the GGFD POM dataset. The dataset better captures the spatiotemporal variability of POC, PON, and their ratio, enabling more consistent characterization of global carbon–nitrogen interactions and marine biogeochemical stoichiometry. Error propagation analysis indicated that uncertainties in PON retrieval and reconstruction dominate the uncertainty in the POC:PON ratio due to its higher variability. Future work should focus on improving PON retrieval accuracy through incorporating additional in-situ observations and higher-resolution physical constraints.

Data availability

The global gap-free POC, PON and POC:PON daily dataset (1998–2023) produced in this study is stored in NetCDF format and openly available at https://doi.org/10.11888/Ocean.tpdc.303488 (Zhang and Liu, 2026).

Author contributions

HL and YZ designed the research. YZ and HL collected the existing in-situ, remote sensing and reanalysis data products. YZ and CL developed the code for the retrieval and reconstruction of POM. YZ and HL performed the analysis. YZ prepared the manuscript with contributions from all co-authors.

Competing interests

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

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

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. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

We would like to thank SeaBASS, the Dryad Digital Repository, and Zenodo for publishing the in-situ POM measurements used in this study, as well as the scientists and crews involved in collecting, processing, and making these data freely available to the scientific community. We also acknowledge the OC-CCI group for providing the satellite data, the Copernicus Marine Service for supplying the Global Ocean Physics Reanalysis data and the reproduced bio-optical products, and NASA OBPG for providing the ocean color satellite products used in this study.

Financial support

This study was funded by Guangdong Major Project of Basic and Applied Basic Research (No. 2023B0303000017), the National Natural Science Foundation of China (No. 42471387 and 42371337), Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011388), the Shenzhen Science and Technology Program (No. JCYJ20230808105709020 and JCYJ20240813142621029), the Disciplines Breakthrough Project in Aerospace Information and Spatiotemporal Intelligence, MOE, China, the National Key Research and Development Program of China (No. 2024YFF0617900), Shenzhen Key Laboratory Program (SYSPG20241211173845013), and Scientific Foundation for Youth Scholars of Shenzhen University (No. 806-000034080293).

References

Aumont, O., van Hulten, M., Roy-Barman, M., Dutay, J. C., Éthé, C., and Gehlen, M.: Variable reactivity of particulate organic matter in a global ocean biogeochemical model, Biogeosciences, 14, 2321-2341, https://doi.org/10.5194/bg-14-2321-2017, 2017.

Barth, A., Alvera-Azcárate, A., Licer, M., and Beckers, J. M.: DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations, Geoscientific Model Development, 13, 1609-1622, https://doi.org/10.5194/gmd-13-1609-2020, 2020.

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

Boyd, P. W. and Trull, T. W.: Understanding the export of biogenic particles in oceanic waters: Is there consensus？, Prog. Oceanogr., 72, 276-312, https://doi.org/10.1016/j.pocean.2006.10.007, 2007.

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., and Weber, T.: Multi-faceted particle pumps drive carbon sequestration in the ocean, Nature, 568, 327-335, https://doi.org/10.1038/s41586-019-1098-2, 2019.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Brewin, R. J. W., Sathyendranath, S., Platt, T., Bouman, H., Ciavatta, S., Dall'Olmo, G., Dingle, J., Groom, S., Jönsson, B., Kostadinov, T. S., Kulk, G., Laine, M., Martínez-Vicente, V., Psarra, S., Raitsos, D. E., Richardson, K., Rio, M. H., Rousseaux, C. S., Salisbury, J., Shutler, J. D., and Walker, P.: Sensing the ocean biological carbon pump from space: A review of capabilities, concepts, research gaps and future developments, Earth-Science Reviews, 217, https://doi.org/10.1016/j.earscirev.2021.103604, 2021.

Canadell, J. G., Ciais, P., Gurney, K., Le Quéré, C., Piao, S., Raupach, M. R., and Sabine, C. L.: An international effort to quantify regional carbon fluxes, Eos, Transactions American Geophysical Union, 92, 81–82, https://doi.org/10.1029/2011EO100001, 2011.

Copinmontegut, C. and Copinmontegut, G.: Stoichiometry of carbon, nitrogen, and phosphorus in marine particulate matter, Deep-Sea Research Part a-Oceanographic Research Papers, 30, 31-46, https://doi.org/10.1016/0198-0149(83)90031-6, 1983.

Demir, K. T., Mathis, M., Kossack, J., Liu, F. F., Daewel, U., Stegert, C., Thomas, H., and Schrum, C.: Variable organic matter stoichiometry enhances the biological drawdown of CO2 in the northwest European shelf seas, Biogeosciences, 22, 2569-2599, https://doi.org/10.5194/bg-22-2569-2025, 2025.

Duan, H. T., Feng, L., Ma, R. H., Zhang, Y. C., and Loiselle, S. A.: Variability of particulate organic carbon in inland waters observed from MODIS Aqua imagery, Environmental Research Letters, 9, https://doi.org/10.1088/1748-9326/9/8/084011, 2014.

Evers-King, H., Martinez-Vicente, V., Brewin, R. J. W., Dall'Olmo, G., Hickman, A. E., Jackson, T., Kostadinov, T. S., Krasemann, H., Loisel, H., Röttgers, R., Roy, S., Stramski, D., Thomalla, S., Platt, T., and Sathyendranath, S.: Validation and Intercomparison of Ocean Color Algorithms for Estimating Particulate Organic Carbon in the Oceans, Front. Mar. Sci., 4, https://doi.org/10.3389/fmars.2015.00251, 2017.

Fagan, A. J., Tanioka, T., Larkin, A. A., Lee, J. A., Garcia, N. S., and Martiny, A. C.: Elemental stoichiometry of particulate organic matter across the Atlantic Ocean, Biogeosciences, 21, 4239-4250, https://doi.org/10.5194/bg-21-4239-2024, 2024.

Falkowski, P. G., Barber, R. T., and Smetacek, V.: Biogeochemical Controls and Feedbacks on Ocean Primary Production, Science, 281, 200-207, 1998.

Fredj, E., Roarty, H., Kohut, J., Lai, J. W., and Ieee: Fast Gap Filling of the coastal ocean surface current in the seas around Taiwan, OCEANS Conference, Shanghai, PEOPLES R CHINA, Apr 10-13, WOS:000386521800094, 2016.

Fumenia, A., Loisel, H., Reynolds, R. A., and Stramski, D.: Relationships between the concentration of particulate organic nitrogen and the inherent optical properties of seawater in oceanic surface waters, Biogeosciences, 22, 2461-2484, https://doi.org/10.5194/bg-22-2461-2025, 2025.

Fumenia, A., Petrenko, A., Loisel, H., Djaoudi, K., DeVerneil, A., and Moutin, T.: Optical proxy for particulate organic nitrogen from BGC-Argo floats, Opt. Express, 28, 21391-21406, https://doi.org/10.1364/oe.395648, 2020.

Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values, Computational Statistics & Data Analysis, 54, 1167-1178, https://doi.org/10.1016/j.csda.2009.09.020, 2010.

Geider, R. J. and La Roche, J.: Redfield revisited:: variability of C:N:P in marine microalgae and its biochemical basis, European Journal of Phycology, 37, 1-17, https://doi.org/10.1017/s0967026201003456, 2002.

Henderson, L. C., English, C. J., Jeng, D. L., Popendorf, K. J., Carlson, C. A., and Close, H. G.: Carbohydrate content controls vertical variations in carbon to nitrogen ratios of organic particles within the euphotic zone in the northwest Sargasso Sea, Communications Earth & Environment, 6, https://doi.org/10.1038/s43247-025-02524-6, 2025.

Jiang, G. J., Loiselle, S. A., Yang, D. T., Gao, C. J., Ma, R. H., Su, W., and Duan, H. T.: An absorption-specific approach to examining dynamics of particulate organic carbon from VIIRS observations in inland and coastal waters, Remote Sens. Environ., 224, 29-43, https://doi.org/10.1016/j.rse.2019.01.032, 2019.

Kharbush, J. J., Close, H. G., Van Mooy, B. A. S., Arnosti, C., Smittenberg, R. H., Le Moigne, F. A. C., Mollenhauer, G., Scholz-Böttcher, B., Obreht, I., Koch, B. P., Becker, K. W., Iversen, M. H., and Mohr, W.: Particulate Organic Carbon Deconstructed: Molecular and Chemical Composition of Particulate Organic Carbon in the Ocean, Front. Mar. Sci., 7, https://doi.org/10.3389/fmars.2020.00518, 2020.

Kim, H., Crow, W., Li, X., Wagner, W., Hahn, S., and Lakshmi, V.: True global error maps for SMAP, SMOS, and ASCAT soil moisture data based on machine learning and triple collocation analysis, Remote Sens. Environ., 298, https://doi.org/10.1016/j.rse.2023.113776, 2023.

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

Le, C. F., Lehrter, J. C., Hu, C. M., MacIntyre, H., and Beck, M. W.: Satellite observation of particulate organic carbon dynamics on the Louisiana continental shelf, J. Geophys. Res.-Oceans, 122, 555-569, https://doi.org/10.1002/2016jc012275, 2017.

Le, C. F., Zhou, X. Y., Hu, C. M., Lee, Z. P., Li, L., and Stramski, D.: A Color-Index-Based Empirical Algorithm for Determining Particulate Organic Carbon Concentration in the Ocean From Satellite Observations, J. Geophys. Res.-Oceans, 123, 7407-7419, https://doi.org/10.1029/2018jc014014, 2018.

Lee, Z., Arnone, R., Hu, C. M., Werdell, P. J., and Lubac, B.: Quantification of uncertainties in remotely derived optical properties of coastal and oceanic waters, Conference on Ocean Sensing and Monitoring II, Orlando, FL, Apr 05-06, WOS:000285624000001, https://doi.org/10.1117/12.849455, 2010.

Liu, D., Bai, Y., He, X. Q., Tao, B. Y., Pan, D. L., Chen, C. T. A., Zhang, L., Xu, Y., and Gong, C. H.: Satellite estimation of particulate organic carbon flux from Changjiang River to the estuary, Remote Sens. Environ., 223, 307-319, https://doi.org/10.1016/j.rse.2019.01.025, 2019.

Liu, H. Z., Li, Q. Q., Bai, Y., Yang, C., Wang, J. J., Zhou, Q. M., Hu, S. B., Shi, T. Z., Liao, X. M., and Wu, G. F.: Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods, Remote Sens. Environ., 256, https://doi.org/10.1016/j.rse.2021.112316, 2021.

Liu, X. M. and Wang, M. H.: Global daily gap-free ocean color products from multi-satellite measurements, International Journal of Applied Earth Observation and Geoinformation, 108, https://doi.org/10.1016/j.jag.2022.102714, 2022.

Martiny, A. C., Vrugt, J. A., and Lomas, M. W.: Concentrations and ratios of particulate organic carbon, nitrogen, and phosphorus in the global ocean, Scientific Data, 1, https://doi.org/10.1038/sdata.2014.48, 2014.

Martiny, A. C., Vrugt, J. A., Primeau, F. W., and Lomas, M. W.: Regional variation in the particulate organic carbon to nitrogen ratio in the surface ocean, Global Biogeochemical Cycles, 27, 723-731, https://doi.org/10.1002/gbc.20061, 2013a.

Martiny, A. C., Pham, C. T. A., Primeau, F. W., Vrugt, J. A., Moore, J. K., Levin, S. A., and Lomas, M. W.: Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter, Nature Geoscience, 6, 279-283, https://doi.org/10.1038/ngeo1757, 2013b.

Matsumoto, K., Rickaby, R., and Tanioka, T.: Carbon Export Buffering and CO₂Drawdown by Flexible Phytoplankton C:N:P Under Glacial Conditions, Paleoceanography and Paleoclimatology, 35, https://doi.org/10.1029/2019pa003823, 2020.

Pahlevan, N., Smith, B., Binding, C., and O'Donnell, D. M.: Spectral band adjustments for remote sensing reflectance spectra in coastal/inland waters, Opt. Express, 25, 28650-28667, https://doi.org/10.1364/oe.25.028650, 2017.

Pauthenet, E., Martinez, E., Gorgues, T., Roussillon, J., Drumetz, L., Fablet, R., and Roux, M.: Contrasted Trends in Chlorophyll-a Satellite Products, Geophysical Research Letters, 51, https://doi.org/10.1029/2024gl108916, 2024.

Redfield, A. C.: On the proportions of organic derivatives in sea water and their relation to the composition of plankton, in: James Johnstone Memorial Volume, edited by: Daniel, R. J., University of Liverpool Press, Liverpool, 176-192, 1934.

Redfield, A. C., Ketchum, B. H., and Richards, F. A., Hill, M. N. (Ed.): The influence of organisms on the composition of sea-water, The composition of seawater: Comparative and descriptive oceanography. The sea: ideas and observations on progress in the study of the seas, Interscience Publishers, New York1963.

Reynolds, R. W. and Smith, T. M.: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation, Journal of Climate, 7, 929-948, https://doi.org/10.1175/1520-0442(1994)007<0929:Igssta>2.0.Co;2, 1994.

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 backscattering coefficient, J. Geophys. Res.-Oceans, 121, 2552-2571, https://doi.org/10.1002/2015jc011408, 2016.

Stramski, D., Joshi, I., and Reynolds, R. A.: Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support of a long-term data record from multiple satellite missions, Remote Sens. Environ., 269, https://doi.org/10.1016/j.rse.2021.112776, 2022.

Stramski, D., Reynolds, R. A., Kahru, M., and Mitchell, B. G.: Estimation of particulate organic carbon in the ocean from satellite remote sensing, Science, 285, 239-242, https://doi.org/10.1126/science.285.5425.239, 1999.

Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M. S., Franz, B. A., and Claustre, H.: Relationships between the surface concentration of particulate organic

https://doi.org/10.5194/essd-2026-453
Preprint. Discussion started: 29 June 2026

carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans, Biogeosciences, 5, 171-201, https://doi.org/10.5194/bg-5-171-2008, 2008.

Tanioka, T., Larkin, A. A., Moreno, A. R., Brock, M. L., Fagan, A. J., Garcia, C. A., Garcia, N. S., Gerace, S. D., Lee, J. N. A., Lomas, M. W., and Martiny, A. C.: Global Ocean Particulate Organic Phosphorus, Carbon, Oxygen for Respiration, and Nitrogen (GO-POPCORN), Scientific Data, 9, https://doi.org/10.1038/s41597-022-01809-1, 2022.

Taucher, J., Boxhammer, T., Bach, L. T., Paul, A. J., Schartau, M., Stange, P., and Riebesell, U.: Changing carbon-to-nitrogen ratios of organic-matter export under ocean acidification, Nature Climate Change, 11, 52-+, https://doi.org/10.1038/s41558-020-00915-5, 2021.

Vantrepotte, V. and Mélin, F.: Temporal variability of 10-year global SeaWiFS time-series of phytoplankton chlorophyll a concentration, Ices Journal of Marine Science, 66, 1547-1556, https://doi.org/10.1093/icesjms/fsp107, 2009.

Wang, C. Y., Zhang, K., Cao, Z. M., Zhou, K. B., Yuan, Z. W., Chen, J. H., Ma, Y. F., Zhou, B., Liu, X., Cai, Y. H., Shi, D. L., and Dai, M. H.: Size-fractionated C:N:P:Si stoichiometry of particulate matter in the subtropical Western North Pacific, Glob. Planet. Change, 246, https://doi.org/10.1016/j.gloplacha.2025.104732, 2025a.

Wang, T. H., Yu, P., Wu, Z. L., Lu, W. F., Liu, X., Li, Q. P., and Huang, B. Q.: Revisiting the Intraseasonal Variability of Chlorophyll-a in the Adjacent Luzon Strait With a New Gap-Filled Remote Sensing Data Set, IEEE Trans. Geosci. Remote Sensing, 60, https://doi.org/10.1109/tgrs.2021.3067646, 2022a.

Wang, Y. Q., Liu, H. Z., and Wu, G. F.: Satellite retrieval of oceanic particulate organic nitrogen concentration, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.943867, 2022b.

Wang, Y. Q., Liu, H. Z., Zhang, Z. X., Wang, Y. R., Zhao, D. M., Zhang, Y., Li, Q. Q., and Wu, G. F.: Ocean Colour Atmospheric Correction for Optically Complex Waters under High Solar Zenith Angles: Facilitating Frequent Diurnal Monitoring and Management, Remote Sens., 16, https://doi.org/10.3390/rs16010183, 2024.

Wang, Y. Q., Liu, H. Z., Wong, C. M., Shen, F., Yu, X. L., Wang, Y. R., Zhang, Y., Zhang, Z. X., Li, Q. Q., and Wu, G. F.: Satellite Retrieval of Water Quality Indicators Under High Solar Zenith Angles, IEEE Trans. Geosci. Remote Sensing, 63, https://doi.org/10.1109/TGRS.2025.3580137, 2025b.

Wasserman, L.: All of statistics: A concise course in statistical inference, 2013.

Wei, J., Yu, X., Lee, Z., Wang, M., and Jiang, L.: Improving low-quality satellite remote sensing reflectance at blue bands over coastal and inland waters, Remote Sensing of Environment, 250, 112029, 2020.

Werdell, P. J. and Bailey, S. W.: An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation, Remote Sens. Environ., 98, 122-140, https://doi.org/10.1016/j.rse.2005.07.001, 2005.

Zhang, C. L., Wang, Y. Y., Bi, R., Sommer, U., Song, G. D., Chen, Z. H., Lin, F., Zhang, J., and Zhao, M. X.: C:N stoichiometry and the fate of organic carbon in ecosystems of the northwest Pacific Ocean, Prog. Oceanogr., 229, https://doi.org/10.1016/j.pocean.2024.103372, 2024a.

Zhang, Y. and Liu, H.: Global gap-free daily 4 km dataset of particulate organic carbon, particulate organic nitrogen and their ratio (1998–2023), National Tibetan Plateau / Third Pole Environment Data Center [dataset], https://doi.org/10.11888/Ocean.tpdc.303488, 2026.

Zhang, Y., Shen, F., Li, R. H., Li, M. Y., Li, Z. X., Chen, S. Y., and Sun, X. R.: AIGD-PFT: the first AI-driven global daily gap-free 4 km phytoplankton functional type data product from 1998 to 2023, Earth Syst. Sci. Data, 16, 4793-4816, https://doi.org/10.5194/essd-16-4793-2024, 2024b.

Zhang, Y., Zhu, P., Xu, G., Liu, C., Wang, Y., Wang, M., and Liu, H.: Satellite Retrieval of Oceanic Particulate Organic Nitrogen Vertical Profiles, Remote Sens., 17, https://doi.org/10.3390/rs17243968 2025.

Zhang, Y., Liu, H. Z., Wang, F., Zhu, P., Zhang, Z. X., Wang, Y. R., Wang, Y. Q., Wu, G. F., and Li, Q. Q.: Toward Applicable Retrieval Models of Oceanic Particulate Organic Nitrogen Concentrations for Multiple Ocean Color Satellite Missions, IEEE Trans. Geosci. Remote Sensing, 62, https://doi.org/10.1109/tgrs.2024.3447699, 2024c.

Zhang, Z., Liu, H., He, X., Zhang, Y., Wang, Y., Wang, Y., Liang, F., Li, Q., and Wu, G.: Satellite retrieval of oceanic particulate organic carbon: Towards an accurate and seamless dataset for the global ocean, The Science of the total environment, 955, 176910, https://doi.org/10.1016/j.scitotenv.2024.176910, 2024d.