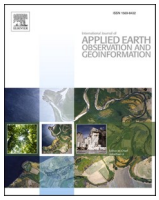
International Journal of Applied Earth Observation and Geoinformation 135 (2024) 104270

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Towards accurate L4 ocean colour products: Interpolating remote sensing reflectance via DINEOF

Christian Marchese [a,b,**], Simone Colella [a], Vittorio Ernesto Brando [a], Maria Laura Zoffoli [a], Gianluca Volpe [a,*]

[a] Institute of Marine Sciences (ISMAR), National Research Council (CNR), Rome, Italy
[b] European Molecular Biology Laboratory (EMBL), Monterotondo, Rome, Italy

ARTICLE INFO

Keywords:
Multivariate DINEOF
Remote sensing reflectance
Chlorophyll-a
Multi spectral Ocean Colour data
Operational oceanography

ABSTRACT

Ocean colour (OC) remote sensing benefits society by providing continuous biological and ecological parameters relevant to sustainable marine resource exploitation. It enhances our understanding of climate change and allows us to monitor oceanographic phenomena over various scales of variability. However, significant data gaps occur daily due to cloud cover, atmospheric correction failures, sun-glint contamination, and satellite coverage limitations. Level 4 (L4) gap-free images are generally created by averaging over specific periods (e.g., weekly, monthly, seasonal) or re-gridding data with coarser resolution to overcome these limitations. These approaches, however, often fail to capture anomalous events or fine-scale resolution processes, calling for more advanced methods. The Data Interpolating Empirical Orthogonal Function (DINEOF) method has proved effective in reconstructing missing OC data and capturing smaller-scale features in noisy fields. To the best of authors knowledge, DINEOF is here used for the first time to interpolate multispectral Remote Sensing Reflectance (Rrs) to produce a consistent and gap-free L4 Rrs dataset, minimizing errors in inferred ocean products, such as Chlorophyll-a (Chl), the most widely used proxy for phytoplankton biomass. Specifically, using a multivariate approach, we assessed the DINEOF technique's capability to reconstruct Rrs, focusing on six bands (412, 443, 490, 510, 555, and 670 nm) and validating the results using extensive *in situ* datasets. Our outcomes show that this "upstream interpolation" method can generate a consistent Rrs dataset, thereby improving the accuracy of L4 Chl predictions when used as input in algorithms for remote Chl estimation. We anticipate further improvements in L4 Rrs accuracy using richer spectral information from upcoming hyperspectral satellite missions. This study highlights the effectiveness of using Rrs as a standalone dataset for DINEOF interpolation. Operationally, it can derivate various gap-free and consistent biogeochemical parameters with reduced uncertainty, thus providing a more reliable and versatile method.

1. Introduction

Remote sensing observations are at the forefront of numerous monitoring systems of the global ocean, such as the Copernicus Marine Environment Monitoring Service (CMEMS) or the National Oceanic and Atmospheric Administration (NOAA) CoastWatch (Liu and Wang, 2018). These observations offer cost-effectiveness and the opportunity for synergistic use of different ocean products. Satellite oceanographic data are increasingly being used to extract crucial marine ecosystem indicators (Blondeau-Patissier et al., 2014; Polovina and Howell, 2005; Racault et al., 2014) and to identify and track biological hotspots (Marchese, 2015; Palacios et al., 2006). Specifically, ocean colour (OC) data are essential to understanding the optical, biological, and ecological aspects of marine ecosystems, providing a valuable means for improving physical and biogeochemical models of the ocean (Yoder et al., 2010). Essential Ocean and Climate Variables (EOVs – ECVs) are environmental parameters that characterize the Earth's climate (Hollmann et al., 2013). These variables include oceanographic parameters like phytoplankton biomass, frequently indexed by the chlorophyll-a concentration (Chl), which can be retrieved from OC data. Consequently, there is an imperative demand for continuous and precise global-scale OC measurements.

* Corresponding author.
** Corresponding author at: Institute of Marine Sciences (ISMAR), National Research Council (CNR), Rome, Italy.
E-mail addresses: christian.marchese@artov.ismar.cnr.it (C. Marchese), gianluca.volpe@cnr.it (G. Volpe).

https://doi.org/10.1016/j.jag.2024.104270
Received 5 August 2024; Received in revised form 28 October 2024; Accepted 9 November 2024
Available online 21 November 2024

Despite the recognized potential of OC data, several constraints limit the volume of usable data, making it challenging to obtain continuous time series. Missing OC data are usually due to clouds obstructing ocean reflectance measurements, the contaminating effects of sun-glint, ice in polar regions, adjacency effects from land, and bottom reflectance in optically shallow waters (Groom et al., 2019; Hilborn and Costa, 2018). These physical limitations lead to satellite data gaps, reducing the satellite's effectiveness. Regarding cloud cover, the seasonal average percentages in the Mediterranean Sea hardly exceed ~ 60 % in winter; the rate decreases in spring, reaching around 35 % in summer and again close to ~ 55 % in autumn (Enriquez-Alonso et al., 2016) The OC data availability can drop dramatically in other marine regions (e.g., the North Atlantic and the Arctic Ocean), where the cloud cover is generally more persistent (Cole et al., 2012) than in the Mediterranean Sea. This can have a limited impact on climate studies, where the scales of variability are much larger. However, it represents a significant drawback when investigating smaller mesoscale phenomena such as those associated with coastal areas. As such, gaps in biogeochemical data retrieved from OC at various space and time scales call for further research into effective data interpolation methods.

While future satellite OC missions and novel analytical tools may reduce data gaps, they still exist in available OC datasets. Traditional methods for filling data gaps in OC satellite images reduce the space–time resolution, while newer methods combine satellite data with numerical models for reconstruction (Konik et al., 2019). Over time, various sophisticated data-driven reconstruction techniques have been devised to effectively address the issue of long-term data gaps. These techniques span from kriging to optimal interpolation, empirical orthogonal function (EOF), and machine learning methods (Ćatipović et al., 2023). Among EOF interpolation methods, the Data INterpolating Empirical Orthogonal Functions (DINEOF; Beckers and Rixen, 2003) is likely the most widely used to fill gaps in OC data time series. DINEOF is a robust and advanced statistical method that identifies the dataset's dominant spatial and temporal patterns, allowing for the accurate reconstruction of missing data without *a priori* statistical information (Taylor et al., 2013).

DINEOF has recently proven effective in filling observational gaps in ocean products derived from satellite data, such as sea surface salinity (Alvera-Azcárate et al., 2016), water turbidity (Alvera-Azcárate et al., 2015), sea surface level (Volpe et al., 2012), and suspended particulate matter (Alvera-Azcárate et al., 2021). It has also been used in a multivariate approach to leverage natural correlations among variables (Alvera-Azcárate et al., 2007). Due to its versatility and robustness, DINEOF has been thus widely applied to remotely sensed data across various marine ecosystems, including the Gulf of Maine (Li and He, 2014), the Gulf of Cadiz (Navarro et al., 2012), and the Mediterranean Sea (Rinaldi et al., 2014). It has also been successfully used in high-latitude regions, such as the Gulf of Alaska (Waite and Mueter, 2013) and the mid-North Atlantic (McGinty et al., 2016). However, sub-polar and polar seas present additional challenges due to significant data gaps caused by frequent cloud cover, sea ice, and low solar angles, which limit valid satellite ocean colour observations. The extreme seasonal cycles in light availability—especially the absence of satellite observations during polar night—further complicate data reconstruction in these regions. Despite these constraints, DINEOF remains a valuable tool for reconstructing missing biogeochemical data in the polar areas. By excluding from the time series days with poor observations and applying masks to avoid interpolation over ice-covered areas, DINEOF can still compensate for limited satellite coverage (Marchese et al., 2017) across these remote and harsh environments. However, it is worth noting that the reduced temporal dimension (i.e., shortened temporal series) may impact the quality of the outcomes of DINEOF for spatially reconstructing satellite datasets (Hilborn and Costa, 2018).

Probably because of their ecological value, sea surface temperature (SST) and Chl (Ćatipović et al., 2023) remain the most widely used variables in studies exploring the biophysical connections in the oceans (Hobday and Hartog, 2014; Volpe et al., 2012) and, as such, are used as input to interpolation procedures. However, the primary product of OC remote sensing is the spectral Remote Sensing Reflectance (Rrs), which is defined as the ratio of the water-leaving radiance to the downwelling irradiance at the sea surface (Morel et al., 1995; Gilerson et al., 2022). Numerous *ad hoc* algorithms exist to retrieve bio-optical properties such as the absorption of coloured dissolved organic matter (a_{CDOM}), particulate backscattering coefficient (b_{bp}), phytoplankton abundance, groups and species, particulate inorganic and organic carbon, and particle size distribution; the totality of these parameters is derived from measured Rrs spectra. Therefore, applying DINEOF techniques directly to primary Rrs instead of ocean products might help derive several gap-free biogeochemical parameters. Furthermore, Rrs measurements typically align more closely with the statistical assumptions behind DINEOF, such as linearity and Gaussian distributions, potentially resulting in more accurate interpolations. However, despite these advantages, to our knowledge, no specific scientific publications directly evaluate the use of DINEOF or any other interpolation approach to Rrs observations. CMEMS currently does not distribute any L4 Rrs products, though this data type could greatly benefit ecological, oceanographic, and climatic applications by providing gap-free radiometric information for deriving ocean variables.

In this context, this study investigated the effectiveness of applying the DINEOF method directly to Rrs data and using the derived Chlorophyll-a (Chl) product as a case study. From a remote sensing perspective, Chl provides an excellent index of phytoplankton biomass. It is associated with the net primary production in marine ecosystems, and physical and biochemical processes strongly influence its variability (Huot et al., 2007). Different analyses have been performed to discriminate among various DINEOF settings and assess their final products' quality.

The findings suggested that the interpolated Rrs can serve as a standalone dataset, simplifying the production of various comprehensive geophysical products without data gaps. The Mediterranean Sea was chosen as a case study because it has relatively low cloud cover compared to other oceanic areas, facilitating the comparative analyses. It, therefore, provided an ideal environment for testing the impact of data density on the interpolation of OC satellite observations.

2. Material and methods

2.1. Datasets

2.1.1. Satellite OC dataset

Two satellite OC datasets were used: the multi-sensor Rrs (sr^{-1}) data and the derived Chl (mg m^{-3}) product. Both were downloaded from the CMEMS web portal as Level 3 (L3) (Product ID: OCEANCOLOUR_MED_BGC_L3_MY_009_143). The L3 data correspond to gridded products with gaps. Briefly, within CMEMS, single-sensor Rrs data are routinely downloaded from the respective space agencies as Level 2 (L2) data, meaning each sensor undergoes its atmospheric correction algorithm depending on the space agency (Volpe et al., 2012; 2019; Colella et al., 2023). The number of sensors contributing to the multi-sensor product time series changed over the years with one sensor (SeaWiFS) in the years 1997 to 2002 to five sensors (MODIS-Aqua, VIIRS-SPP/NOAA, and OLCI-Sentinel3A/B), at the time of writing. After a series of processing steps aimed at addressing sensor-specific issues and resampling at 1 km spatial resolution, single-sensor Rrs are merged into a multi-sensor product (i.e., Rrs_{L3}) that accounts for inter-sensor biases (Volpe et al., 2019; Colella et al., 2023). Therefore, starting from the multi-sensor Rrs_{L3}, the generation of the Chl product (Chl_{L3}, Fig. 1a) involves four main steps:

(1) Applying the MedOC4.2020 regional Chl algorithm, suited for Case I waters (Colella et al., 2023), to the entire Rrs_{L3} field. Note that the MedOC4.2020 shares the functional form of traditional

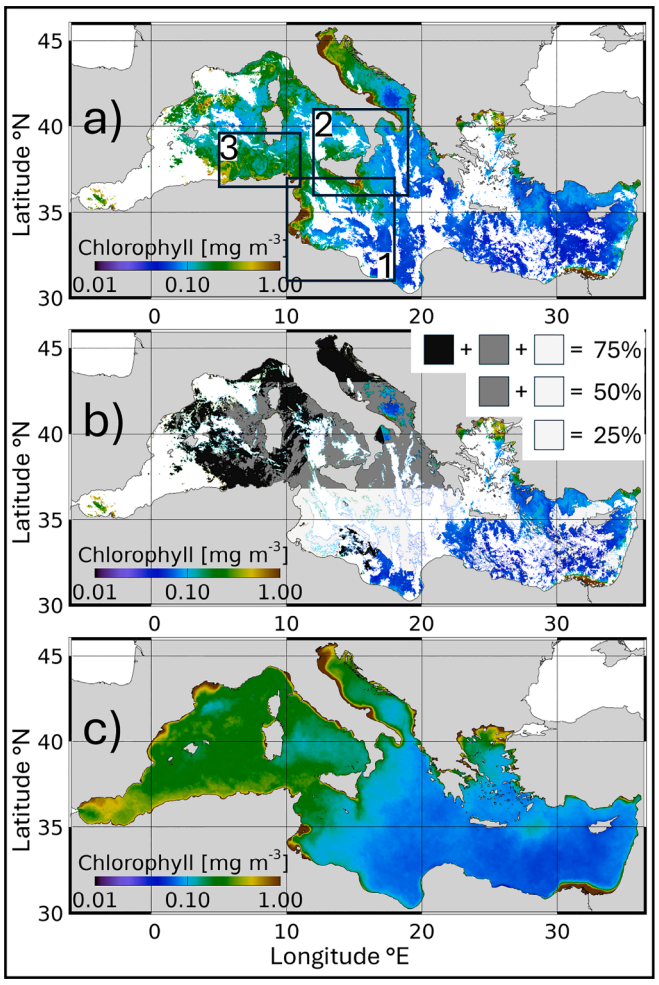


Fig. 1. a) Original L3 Chl observations for February 15th, 2023, as available from CMEMS multi-sensor product time series. The three numbered areas were used to evaluate the spatial coherency of the interpolated products. b) An example of artificial cloud cover, with pixels marked in light grey, corresponding to 25 % of cloud cover, dark grey to 50 %, and black to 75 %. c) An example of daily Chl climatology for February 15th was used to build the input data matrix and in the validation analysis.

empirical OCx NASA algorithms (O'Reilly et al., 2000) but features polynomial coefficients optimized for the Mediterranean Sea (Volpe et al., 2007);

(2) Applying the AD4 regional Chl algorithm, suited for Case II waters (Berthon and Zibordi, 2004), to the entire Rrs_{L3} field;

(3) determination of the pixels belonging to the two water types and identifying those that do not strictly belong to either type. Specifically, identifying Case I or Case II water type membership involves comparing the satellite Rrs_{L3} spectra (from 443 nm to 670 nm) with *in situ* Rrs spectra from the MedBiOp dataset (see section 2.1.2) at pixel scale and using k-means clustering to distinguish pure Case II spectra from others (Colella et al., 2023);

(4) merging of the two Chl images according to the water type membership of each pixel.

In the Rrs_{L3} and Chl_{L3} images used in the analysis, gaps were introduced to simulate 25 %, 50 %, and 75 % of cloud coverage (Fig. 1b, see section 2.3). Specifically, data gaps were progressively added for each daily image from the same day of previous years until they reached the defined cloud cover threshold. The cloud patterns were sourced from other images to ensure the coverage looks authentic because clouds are not homogeneously distributed.

Finally, climatology daily maps were created using the original L3 data from 1998 to 2023, falling into a moving temporal window of ± 5 days (Fig. 1c). In other words, the climatological value for each calendar day was computed by averaging the variable (either single Rrs_{L3} bands or Chl_{L3}) from the same date across years within an eleven-day window, encompassing the five days before and after the reference date. These products were referred to as Rrs_{Clima} and Chl_{Clima} according to the Rrs or Chl data. Table 1 summarizes the key parameters adopted in this study.

2.1.2. *In situ* OC dataset

The MedBiOp dataset (Volpe et al., 2019) was extended to 2023 (Volpe et al., 2019; Colella et al., 2023) to be used as a reference to compute the uncertainties associated with satellite interpolated products via matchup analysis. The MedBiOp dataset includes optical and Chl data, whose space–time distribution is shown in Fig. 2. The validation dataset includes additional radiometric measurements from three fixed stations (marked with stars in Fig. 2). These stations include two AERONET sites, which are equipped with above-water hyperspectral radiometers (AAOT and Casa Blanca located in the North Adriatic Sea and off the north-eastern coast of Spain, respectively), and the Meda of Lampedusa Island in the south-central basin equipped with in-water hyperspectral radiometers (Cazzaniga and Mélin, 2024; Di Sarra et al., 2019; Zibordi et al., 2021). AAOT has provided data since 2002, while Casa Blanca and the Meda of Lampedusa have only provided data since 2019. In these three fixed stations, Chl measurements were only sporadically available.

2.2. DINEOF interpolation procedure

Missing data reconstruction was carried out through the DINEOF method, initially introduced by Beckers and Rixen (2003) and successfully applied in the Mediterranean Sea by Volpe et al. (2018, 2012). The DINEOF temporal interpolation works sequentially. Firstly, in our case, a data matrix is built using nine days of satellite L3 data: four before and four following the day that must be interpolated (Fig. 3). This configuration of the input data matrix accounts for the short decorrelation time scale of OC data over the Mediterranean Sea (Volpe et al., 2018). An analysis was also conducted using a larger time window of ±7 days that did not produce significantly different results (results not shown). Therefore, choosing a nine-day window struck an optimal trade-off between capturing as many relevant images as possible and minimizing the

Table 1
List of symbols used throughout this work. In the Chl_{insitu}, the equation is Z_{pd} the light penetration depth (m) (roughly equivalent to one-fifth of the euphotic depth), and K_d is the diffuse light attenuation coefficient (m^{-1}) (Volpe et al., 2007).

Symbol	Description
Chl_{L3}	Non-interpolated (L3) satellite Chl
Rrs_{L3}	Non-interpolated (L3) satellite Rrs observations
Chl_{Clima}	Mean climatological Chl values
Rrs_{Clima}	Mean climatological Rrs values
Chl_{L4}^{Chl}	Gap-free (L4) Chl concentrations derived from the L3 Chl processing over L3 Chl observations
Chl_{L4}^{sRrs}	Gap-free (L4) Chl derived from the application of the Chl algorithm (MedOC4.2020) after the single-band interpolation processing of the L3 Rrs (one run per Rrs band)
Chl_{L4}^{mRrs}	Gap-free (L4) Chl derived from the application of the Chl algorithm (MedOC4.2020) after the simultaneous multi-band interpolation processing of the L3 Rrs (one run per Rrs spectrum)
Chl_{insitu}	$\frac{\int_0^{Z_{pd}} Chl(z)e^{-2\bullet K_d \bullet Z_{pd}}dz}{\int_0^{Z_{pd}} e^{-2\bullet K_d \bullet Z_{pd}}dz}$
Rrs_{L4}^{sRrs}	Gap-free (L4) Rrs derived from the single-band interpolation processing of the L3 Rrs (one run per Rrs band)
Rrs_{L4}^{mRrs}	Gap-free (L4) Chl derived from the simultaneous multi-band interpolation processing of the L3 Rrs (one run per Rrs spectrum)

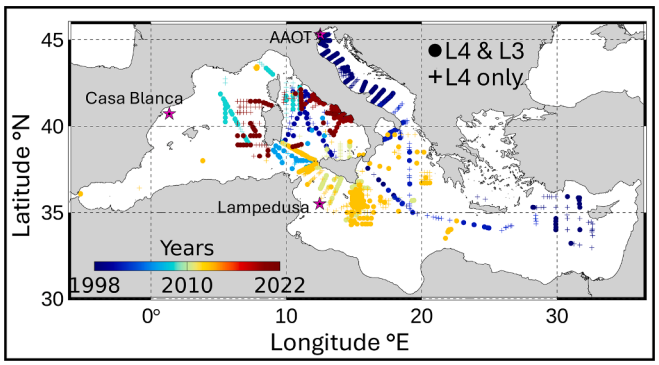


Fig. 2. Space-time distribution of the satellite-*in situ* matchup dataset with crosses denoting missing L3 satellite observations and filled circles indicating the location of matching satellite-*in situ* measurements. The pink stars indicate the position of fixed stations (Casa Blanca, Lampedusa and AAOT). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

impact of unrelated observations on the reconstruction process. Secondly, all missing data in the matrix are replaced with respective daily climatological mean values (Rrs_{Clima} or Chl_{Clima}, see Table 1 for naming convention). A binary mask, namely "holes", is built with zeros corresponding to effective observations and ones where climatological mean values were used. The replacement is made through a smoothing procedure that prevents the creation of artificial gradients.

Thus, the resulting data matrix constitutes the input to the iterative EOF procedure, which uses the singular-value decomposition (SVD) technique (Taylor et al., 2013). After each iteration, the input data matrix to the following SVD iteration is built with original observations in correspondence of the zeros within the holes mask and with the field reconstructed from the SVD output of the previous iteration for holes equal to 1. The reconstruction uses the number of modes corresponding to the iteration number. After the first iteration, the climatology is replaced with the first EOF mode, which is subsequently replaced with the field reconstructed with the first two modes after the second iteration, and so on. DINEOF explicitly identifies areas with different sources of space–time variability through iterative EOF estimation. To determine the optimal number of iterations, the variance explained by each mode is compared with the one explained by the noise, which is, in turn, determined by an independent EOF run performed using a random number matrix with the exact dimensions as the input data matrix. The number of modes may thus differ daily. This short data temporal variability associated with the input data matrix (only nine days) called for a test on the validity of the abovementioned approach used to define the optimal number of modes. We observed that all available (nine) modes did not yield significantly better results than those obtained with the number of modes defined by noise variance (results not shown), but using all modes doubled the computational time. The final interpolated field is a multi-scale product. It is made of original L3 observations in correspondence with holes = 0 and of interpolated data in correspondence with original data gaps (holes = 1). The two fields are merged through a smoothing procedure to prevent spurious spatial gradients.

2.3. DINEOF application

DINEOF was applied to each daily image with at least one *in situ* measurement for 3910 daily satellite images over 25 years. This processing was performed using the best configuration outcome revealed by a previous analysis. Twelve L3 Rrs satellite images were processed under varying synthetically applied cloud cover conditions (25 %, 50 %, and 75 %) to original L3 Rrs and Chl images. The choice to interpolate twelve days (the 15th of every month of 2023) spread over one year on one side ensured that the data captured a sufficiently large environmental variability and, on the other, helped minimize computational time and resources. For assessing the accuracy and efficiency of applying the DINEOF interpolation method directly to Rrs *versus* its application to Chl, three different runs allowed testing the "upstream" versus the "downstream" interpolation approaches. In the first run (hereafter referred to as L3 Chl Processing or downstream interpolation), the DINEOF interpolation was directly applied to the original Chl_{L3} observations, estimated from the non-interpolated Rrs (Rrs_{L3}). This processing resulted in the gap-free (L4) Chl referred hereafter to as Chl_{L4}^{Chl}.

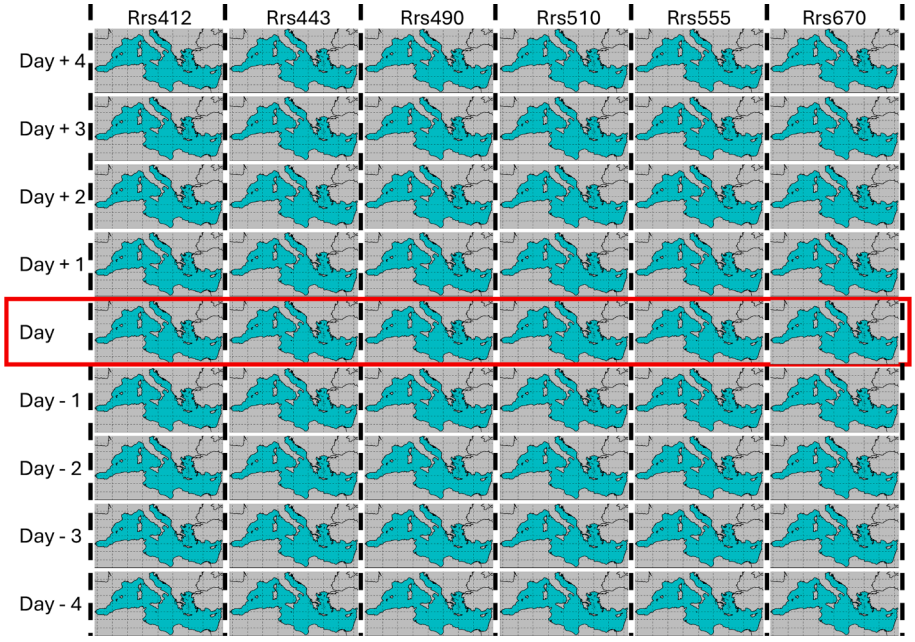


Fig. 3. Schematics of the input data matrix with the bands on the columns and days on the rows, the day that must be interpolated (red rectangle), plus and minus four days. In the single-band approach, the bands are interpolated via individual runs – one run per column – while in the multi-band approach, all bands are used in one single run. In the current work, the number of sea pixels – green areas. is 1957817. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Accuracy metrics used throughout this work to compare the estimated (satellite) dataset, X_i^E to a reference, measured in-situ dataset X_i^M. $\overline{X}^E$ and $\overline{X}^M$ are the average fields of both estimated and reference datasets, respectively. N corresponds to the sampling size. Apart from RPD (mean relative percent difference) and APD (mean absolute percent difference), when applied to Chl data, all other metrics are computed over the log10-transformed data. The S (slope) and I (intercept) are the type-2 linear regression coefficients.

Mathematical expressions of the accuracy metrics
$S = \frac{\sum_{i=1}^{N}(X_i^E-\overline{X}^E)^2-\sum_{i=1}^{N}(X_i^M-\overline{X}^M)^2+\left\{\left[\sum_{i=1}^{N}(X_i^E-\overline{X}^E)^2-\sum_{i=1}^{N}(X_i^M-\overline{X}^M)^2\right]^2+4\left[\sum_{i=1}^{N}(X_i^E-\overline{X}^E)(X_i^M-\overline{X}^M)\right]^2\right\}^{1/2}}{2\sum_{i=1}^{N}(X_i^E-\overline{X}^E)(X_i^M-\overline{X}^M)}$
$I = \overline{X}^E - S\overline{X}^M$
$r^2 = \frac{\left[\sum_{i=1}^{N}(X_i^E-\overline{X}^E)(X_i^M-\overline{X}^M)\right]^2}{\sum_{i=1}^{N}(X_i^E-\overline{X}^E)^2\sum_{i=1}^{N}(X_i^M-\overline{X}^M)^2}$
$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i^E-X_i^M)^2}{N}}$
$bias = \frac{1}{N}\sum_{i=1}^{N}(X_i^E-X_i^M)$
$RPD = 100 \bullet \frac{1}{N}\sum_{i=1}^{N}\frac{X_i^E-X_i^M}{X_i^M}$
$APD = 100 \bullet \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert X_i^E-X_i^M\rvert}{X_i^M}$

Therefore, this procedure embodies the traditional method of filling in missing data for satellite ocean products that are usually obtained at the concluding stages of a workflow, frequently involving specific algorithms (e.g., Volpe et al., 2007; Liu and Wang, 2022, 2019).

The two other runs focused on directly interpolating the Rrs_{L3} field (i.e., upstream interpolation). In the Single Band Processing, each spectral band was interpolated individually as a standalone process using the nine-day data of the same band hereafter referred to as Rrs_{L4}^{sRrs}. The latter was afterward used as input for the Mediterranean Chl estimation using the described regional process (see section 2.1.1) yielding the gap-free (L4) Chl derived from the Rrs_{L4}^{sRrs} and was referred to Chl_{L4}^{sRrs}. In the Multi Band Processing, the input data matrix was built using all bands for the nine days simultaneously, thus interpolating the Rrs_{L3} spectrum *to retrieve* Rrs_{L4}^{mRrs}. Since all bands were simultaneously used, this approach can be considered a multivariate use of DINEOF (Alvera-Azcárate et al., 2007). After interpolation, the *retrieved* Rrs_{L4}^{mRrs}. dataset was used to compute the gap-free (L4) Chl, Chl_{L4}^{mRrs}, using the same regional process.

2.4. DINEOF performance analysis and statistical metrics

Here, to qualify the DINEOF interpolation results, we performed two analyses meant to characterize the final output statistically: 1) assessment of satellite L4 against satellite L3 data (satellite Chl product comparison, see section 3.1) and 2) evaluation of satellite data against *in situ* observations (matchup analysis, see section 3.2).

In the first analysis (i.e., the satellite Chl product comparison), several spatially coherent valid Chl_{L3} pixels (synthetic clouds) were set aside. These pixels were not used in the interpolation procedure and were considered references for comparing the goodness of the interpolated fields. Note that for each interpolation run (i.e., each L4 Chl product), three interpolated results were available from the 25 %, 50 %, and 75 % cloud coverage of satellite images. This provided an estimate of the impact of the total amount of available observations over the goodness of the results, allowing us to evaluate the limit of applicability of the interpolation procedure. In other words, it helped determine the expected uncertainty associated with a given interpolated field as a function of the number of available observations. Besides providing many observations in the final comparison, this exercise also allowed the visual control of any spurious structure resulting from the interpolation procedure.

The *in situ* data set was used in the matchup analysis to assess the pixel-scale satellite data regarding basic statistical parameters, as reported in Table 2. For this analysis, satellite Rrs_{L3}, Chl_{L3}, Chl_{Clima}, and all the L4 interpolated products were extracted from a 3x3 grid pixels centred over the geographical coordinates of *in situ* measurements. At least 50 % of the 3x3 grid pixels for the matchup station must be valid. The variation coefficient (the ratio between the standard deviation and the mean value) computed over the 3x3 pixel window must be lower than 20 %. None of the *in situ* data in the matchup exercise was part of the calibration dataset used to tune the MedOC4.2020 and AD4 Chl algorithms. Since the processed interpolated data were gap-free daily fields, all *in situ* observations within the solar day (24 h) constituted potential valid matchups without any further temporal constraint. After this screening, the median value of the 3x3 pixel grid was contrasted against co-located *in situ* measurement. These criteria yield more than 4600 and roughly 2000 matchups for Rrs and Chl, respectively (Table 3). Other criteria allow for discerning the impact of the interpolation procedure on the overall quality of the final product, which is worth remembering. It comprises available original observations and interpolated data where the observations were missing. For example, the statistics associated with comparing the original L3 and *in situ* observations somehow define the best target of agreement that the L4 – *in situ* comparison should tend to. Similarly, comparing only interpolated

Table 3
Number of *in situ* – satellite matchup points as a function of the dataset (Chl or Rrs) and processing level (observations, L3, interpolated, L4). The data used in the matchup exercises do not include those used to develop the Chl algorithms.

Dataset	Processing Level	Criteria	# of Matchups
Chl	L3	**All**	**821**
	L4	Without L3	1180
		All	**2001**
Rrs	L3	In-water radiometry only	552
		All	**3518**
	L4	In-water radiometry only	855
		In-water radiometry only without L3	300
		without L3	1125
		All	**4643**

pixels with *in situ* observations provides a means to assess the role of the interpolation procedure over the final product.

3. Results and discussion

3.1. Satellite chlorophyll product comparison

To evaluate the accuracy of the L4 Chl products generated by the DINEOF interpolation method through statistical parameters (Table 2), we compared L4 fields over artificially gapped pixels against Chl_{L3} observations (similar to Hilborn and Costa, 2018). We also compared Chl_{L3} against Chl_{Clima}, where no interpolation method was applied, assuming that it might be the only available source for inferring the daily Chl values (Fig. 4). Indeed, Chl_{Clima} values are also used as a first guess within our DINEOF interpolation processing (see section 2.2). Chl_{Clima} showed a tendency to overestimate Chl_{L3}, as evidenced by the positive bias of 0.046 mg m^{-3}, and RPD of 16 %. While it is impressive that it can explain as much as 65 % (r^2) of the daily chlorophyll variability at the km spatial scale, it remains a poor predictor of the daily Chl field, particularly in highly variable marine environments, where extreme events such as Marine Heat Waves (MHW) can impact on phytoplankton production and community composition (Li et al., 2024; Arteaga and Rousseaux, 2023). Overall, climatology or re-gridding data into a coarser grid can help increase data coverage and thus allow for consistent monitoring and analysis over time. However, these basic methods may not be sufficient to capture specific anomalous bloom events or variations during the missing periods, making it challenging to interpret correctly, for example, phytoplankton dynamics in time and space (Cole et al., 2012).

The outcomes associated with applying the DINEOF method to both L3 Chl and L3 Rrs are presented in Fig. 5. Predictably, a slight decrease in accuracy was observed with increasing cloud cover, significantly when the 50 % threshold was surpassed (Alvera-Azcárate et al., 2005). Overall, as cloud cover increases from 25 % to 75 %, there is a noticeable rise in APD from 9 % to 19 %, with RMSE and bias increasing to 0.122 mg m^{-3} and 0.024 mg m^{-3}, respectively (Fig. 5). However, this analysis's first essential and significant outcome is that all DINEOF-derived fields (Chl_{L4}^{Chl}, Chl_{L4}^{sRrs} and Chl_{L4}^{mRrs}) do perform better than Chl_{clima} under all cloud cover scenarios (Fig. 5). This result is highlighted by the statistical metrics (higher r^2 and lower RMSE and RPD) that suggest greater accuracy and consistency when using DINEOF approaches. Even in the worst scenario with 75 % clouds, the Chl_{L4}^{Chl} product (Fig. 5g) showed an increase of approximately 6 % in r^2, while reductions of ~ 26 % in RMSE and ~ 43 % in RPD when comparing to Chl_{clima} (Fig. 4). This result indicates that the DINEOF algorithm successfully assimilates the variability of the initial data matrix and accurately incorporates it into the resulting output. Indeed, r^2 values of 0.69 or higher for the various L4 Chl products derived from DINEOF under the three different cloud cover conditions (Fig. 5) somewhat underscore the reliability of the interpolation method. DINEOF allows a more accurate reconstruction of missing data by identifying the dominant spatial and temporal patterns (Alvera-Azcárate et al., 2005). Instead, filling gaps using climatological values may suppress the natural inter-annual variability. DINEOF primarily retains information at meso and large scales (Liu and Wang, 2018). Still, it can also be applied to high spatial resolution biogeochemical data to resolve complex coastal dynamics (Alvera-Azcárate et al., 2024).

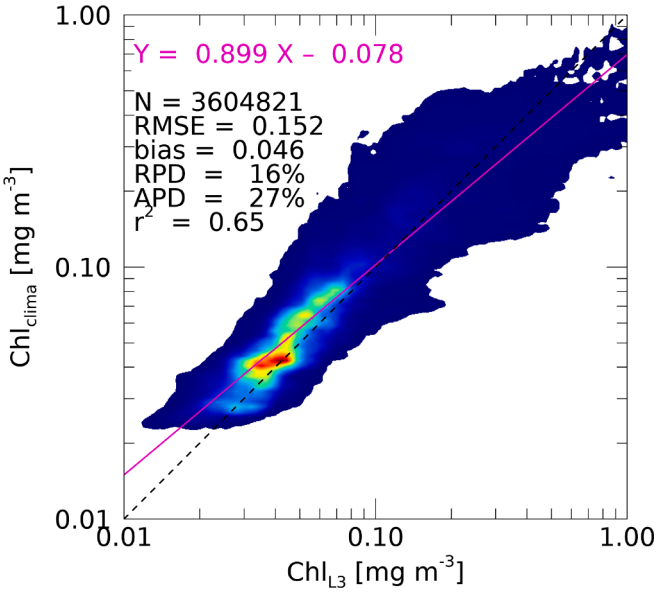


Fig. 4. Scatterplot of the Chl climatology against original satellite L3 observations in correspondence with the 25% cloud cover pixels (see main text for more details).

Interestingly, Chl_{L4}^{Chl} and Chl_{L4}^{sRrs} exhibited very similar accuracy within the different cloud cover percentages (Fig. 5 – first and second columns). This lack of differences suggests that using the DINEOF algorithm to individually interpolate the six Rrs_{L3} spectral bands to create a gap-free dataset, from which retrieve Chl_{L4}^{sRrs}, produces similar results than directly interpolating the original Chl_{L3} observations. However, to provide a comprehensive and consistent data archive of gap-free OC products (Kd490, CDOM concentration, IOPs, etc.), interpolating the Rrs is still preferable to interpolating the single parameters. On the contrary, Chl_{L4}^{mRrs} (Fig. 5 – the rightmost column) yielded higher accuracy across all levels of cloud cover. For instance, within 75 % of cloud cover Chl_{L4}^{mRrs} (Fig. 5i) showed a higher r^2 value of 0.74 than Chl_{L4}^{Chl} and Chl_{L4}^{sRrs} at 25 % (Fig. 5a and b) and similar RMSE, bias, and RPD. This last outcome clearly shows that interpolating the six bands (i.e., 412, 443, 490, 510, 555, and 670 nm) simultaneously in a single run via a multivariate DINEOF approach allows producing a consistent and gap-free Rrs field (i.e., Rrs_{L4}^{mRrs}), which is here used as input of the Mediterranean-specific regional Chl algorithm.

The performance improvement of multi-band processing arises from the increased amount of relevant information, which is only accessible through this method, unlike others such as L3 Chl Processing or Single Band Processing. Using more Rrs bands enhances spatiotemporal coverage. By incorporating all available spectral information, DINEOF more accurately identifies and reproduces the principal modes of variability in the dataset (Alvera-Azcárate et al., 2021). In our case, the multivariate approach captures more modes of variability by exploiting the closer relationships between different Rrs bands (Huot and Antoine, 2016). In other words , the multivariate version of DINEOF, when handling multiple Rrs bands simultaneously, allows for more sophisticated reconstructions through EOF decomposition, which improves gap-filling by leveraging correlations between the different spectral bands. As such, the algorithm can capture subtle spectral correlations and variations that might have been missed when processing each Rrs band individually. Moreover, correlations between variables can also help regions where one variable has more complete coverage than others, resulting in better overall reconstructions than a univariate approach (Alvera-Azcárate et al., 2007). For example, missing Chl data can be reconstructed using correlations with sea surface temperature or other available variables (Alvera-Azcárate et al., 2007). This aspect could be relevant in polar and sub-polar seas with prolonged high cloud cover extent (Bélanger et al., 2013). Thus, more significant input information provides a more comprehensive representation of the data's natural variability and recurring patterns, leading to more reliable and accurate reconstructions of missing values. In this regard, we also repeated the multivariate interpolation at 25 % of cloud cover, feeding DINEOF with Rrs_{L3} at only five spectral bands (i.e., those used to compute the Chl: 443, 490, 510, 555, and 670 nm). We observed very similar results when

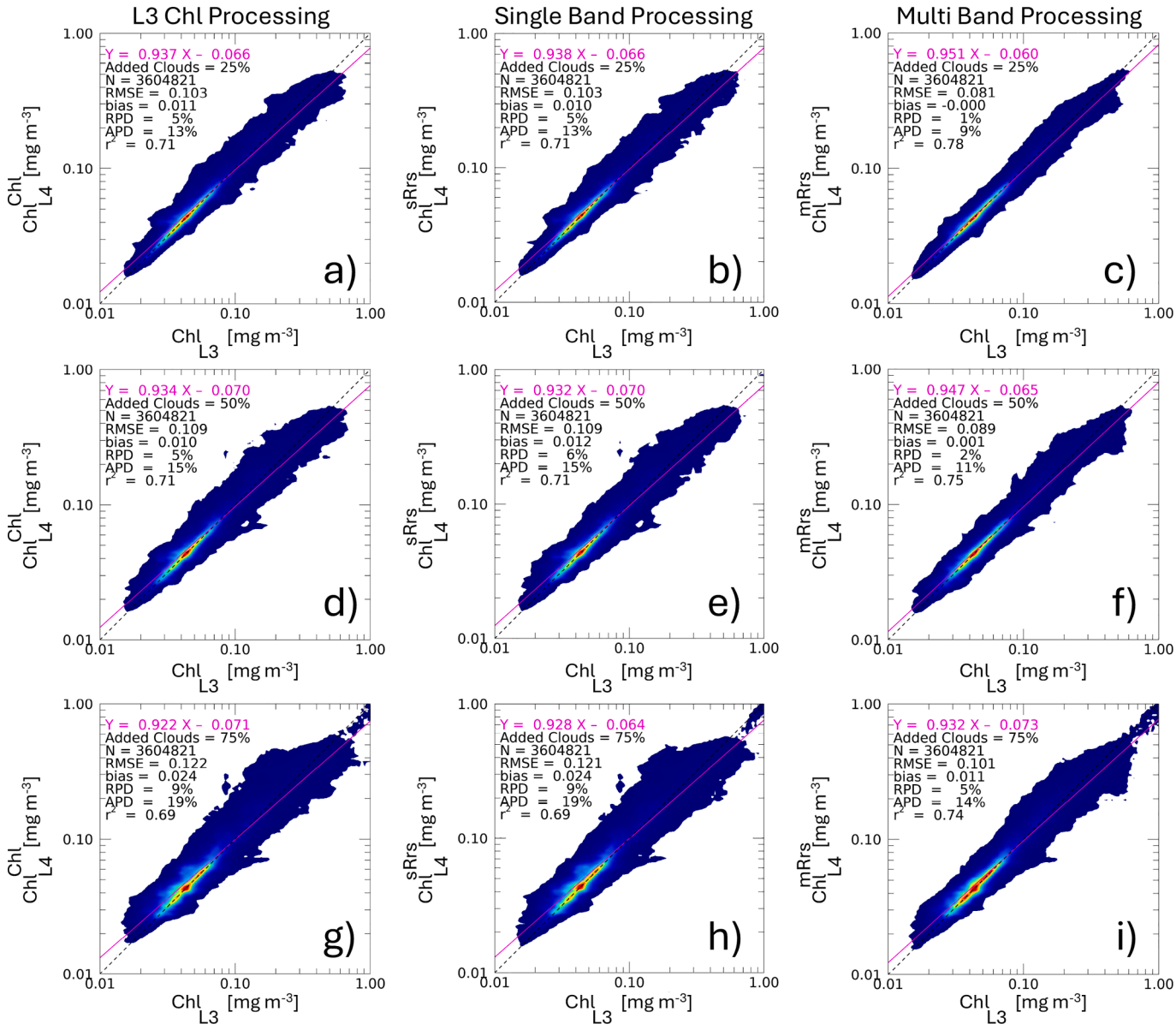


Fig. 5. Scatterplot of the L4 Chl against original satellite *Chl*!“ observations in correspondence of the 25 % (a, b, c), 50 % (d, e, f), and 75 % (g, h, i) cloud cover pixels. The L4 outputs are derived from interpolation of *Chl*!” (*Chl*!# $%&,in panels a, d, and g,), from single-band Rrs interpolation processing (*Chl*!# '()', in panels b, e and h) and multi-bandRrs interpolation processing (*Chl*!# *()', in panels g, f and i). To facilitate the comparison, statistics are computed over the same set of pixels (N = 3604821).

comparing the statistics of this last run (see Figure S1 in the supplementary material) with those in Fig. 5c. However, the interpolation using all six bands concurrently (Fig. 5c) indicated a marginally better result. Therefore, we speculate that when the number of interpolated spectral bands via the multivariate approach is more significant, more conspicuous differences may become apparent compared to the univariate approach, mainly because of the multivariate nature of the Rrs and correlations between wavelengths (Huot and Antoine, 2016). Based on this result and given that the multi-band interpolation outperformed the single-band interpolation, we foresee that increasing spectral resolution in the region 412 – 670 nm (i.e., using hyperspectral data) could result in more accurate and reliable satellite Rrs gap-free datasets for bio-optical parameter retrieval. However, this is a non-trivial task, and further research is needed to test the multivariate DINEOF method for handling hyperspectral data to balance the final output performance with the computational complexity. This may be possible in the foreseeable future thanks to new missions such as NASA's PACE (Plankton, Aerosol, Cloud, ocean Ecosystem). From a computational standpoint, reconstructing multiple variables together should be more efficient than running separate reconstructions for each variable. Incorporating hyperspectral data into the multivariate approach holds promise for achieving more accurate and consistent reconstructions across the entire spectral range, enhancing spatial and spectral data quality without compromising computational efficiency.

Comparing interpolation performance under various cloud cover conditions is crucial to evaluate their reliability and effectiveness in mapping spatially continuous fields for ecosystem modelling (Gregg, 2008), analysing phytoplankton phenology and bio-physical correlations (Marchese et al., 2019, 2017; Mayot et al., 2020), and partitioning the ocean surface (Huot et al., 2019; Marchese et al., 2022). In general, areas with high cloud coverage experience more significant errors in the reconstruction (Alvera-Azcárate et al., 2011). When visually checking maps (Figure S2 in supplementary material), all processing methods effectively compensate for the cloud cover to some extent, maintaining

detail and continuity in the Chl data presentation. However, zooming in on specific areas (Fig. 1a) reveals subtle differences among the various processing approaches and imposed cloud percentages (Fig. 6). Indeed, all methods consistently showed patchy behaviour with local Chl underestimation/overestimation within the three cloud cover percentages (i.e., 25 %, 50 %, 75 %). However, upon closer examination, Chl_{Clima} exhibited a higher magnitude in the differences, both positive and negative, than the DINEOF-derived fields, reflecting its inability to predict Chl within diverse zones (Fig. 6a, 6e, and 6i). This is particularly noticeable at the 25 % and 50 % levels but less evident at 75 %. The lowest differences were displayed by Chl_{L4}^{mRrs} across all levels of cloud cover, demonstrating better performance with areas of overestimation and underestimation less marked (Fig. 6d, h, and l). Overall, the differences were more balanced and presented fewer extreme values than the other methods.

Fig. 6 shows that the multi-band processing interpolation is more accurate than the other tested approaches. Nevertheless the DINEOF reconstruction are impacted by several factors, including data quality and the number of available observations in the time series. For instance, the accuracy of the input data is heavily influenced by atmospheric correction, which is critical for ensuring reliable ocean color retrievals. Therefore, effective preprocessing, including robust atmospheric correction and quality control, is essential to minimize errors in DINEOF reconstructions (Hilborn and Costa, 2018). Additionally, underrepresented oceanic processes are more challenging to reconstruct, making data coverage crucial. The total amount and distribution pattern of the data availability can vary across ocean basins and over the years thus influencing the interpolation outcomes. For instance, Zhao et al. (2024) found that in the oligotrophic regions between 40°N and 40°S, daily Chl coverage can be improved by an average of 43 %–55 %, resulting in an almost complete daily time series. In contrast, the polar oceans saw only a 4.66 % increase in coverage. Therefore, greater spatial and temporal availability improves reconstruction accuracy by allowing more EOF modes to be calculated, capturing finer-scale variability in the data (Hilborn and Costa, 2018).

3.2. Matchups analysis

The primary objective of the previous section was to define the best interpolation configuration, and the outcome was that the multi-band processing consistently outperformed all other tested approaches. Based on this result, this section aimed to evaluate Satellite L3 and L4 Rrs and Chl products resulting from the multi-band processing by comparing them to *in situ* measurements. Validation is crucial for remote sensing, as it quantitatively assesses whether satellite observations meet user requirements and agency expectations and thus suit their intended use (Concha et al., 2021).

Overall, the Rrs_{L3} observations agreed well with the *in situ* Rrs measurements across all bands (Fig. 7). However, 412 and 670 nm bands exhibit higher RMSE than mid-spectrum bands (i.e., 443 nm, 490 nm, 510 nm, and 555 nm). While Rrs at 412 nm is generally underestimated, at 670 nm, it is slightly overestimated. These trends are consistent across both observed Rrs_{L3}, and interpolated Rrs_{L4}^{mRrs} data. However, the observed data typically exhibited better performance. Nonetheless, the spectral bands between 443 and 555 nm showed an excellent performance in both observed and interpolated data, with RPD between 12 and 29 % (Fig. 7). These central bands are critical for Chl retrieval regardless of the algorithms used. In contrast, the band at 670 nm here gives only marginal information to identify the optical water type membership. These results demonstrate that the DINEOF interpolation with the multi-band approach can provide robust gap-free Rrs data suitable for distribution by CMEMS as an L4 product.

A good agreement was found when comparing Chl_{L3} satellite observations to Chl_{insitu}, (Fig. 8a) with an RMSE = 0.26, $r^2 = 0.79$, and a bias equal to 0.0053, indicating excellent performance of the regional algorithms when applied to Rrs_{L3} data. Furthermore, this first comparison

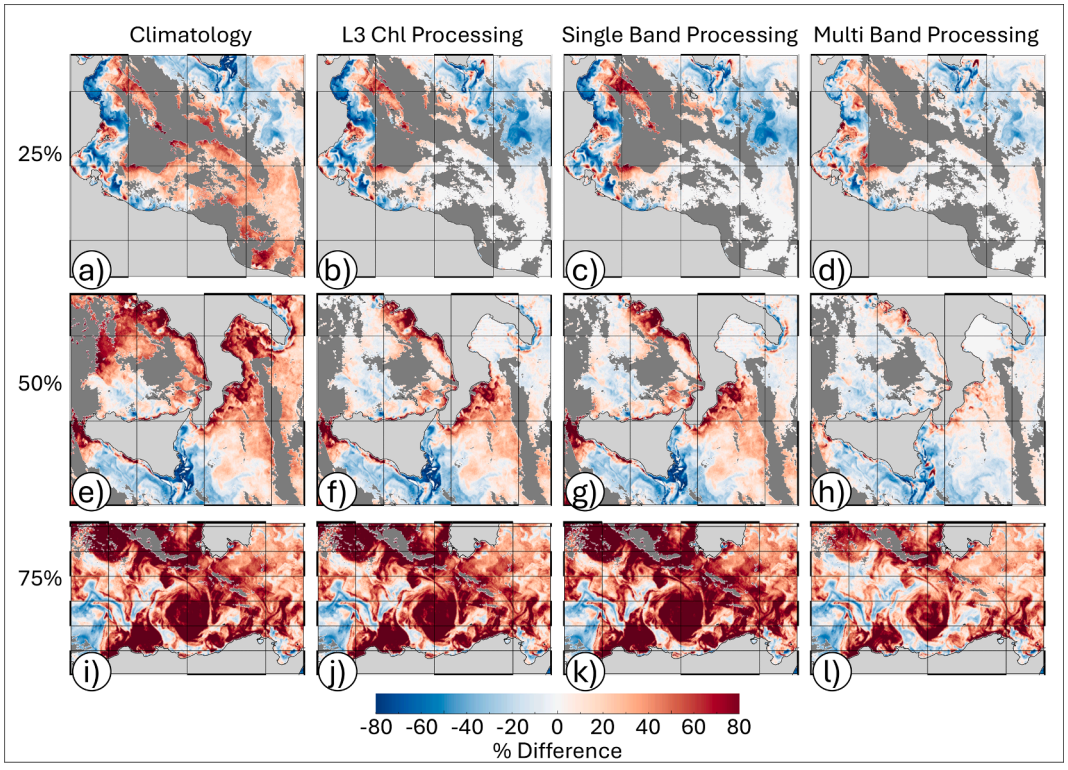


Fig. 6. Example of the percent difference between DINEOF-derived fields according to the three column's processing configurations and original *Chl*!" observations in correspondence of the boxes defined in Fig. 1a for the 15th February 2023. Each box was used to qualitatively verify the interpolation process under the different cloud cover percentages (see Section 2.3); these boxes provided a closer zoom on the various fields, which might otherwise have appeared too similar (see Figure S2).

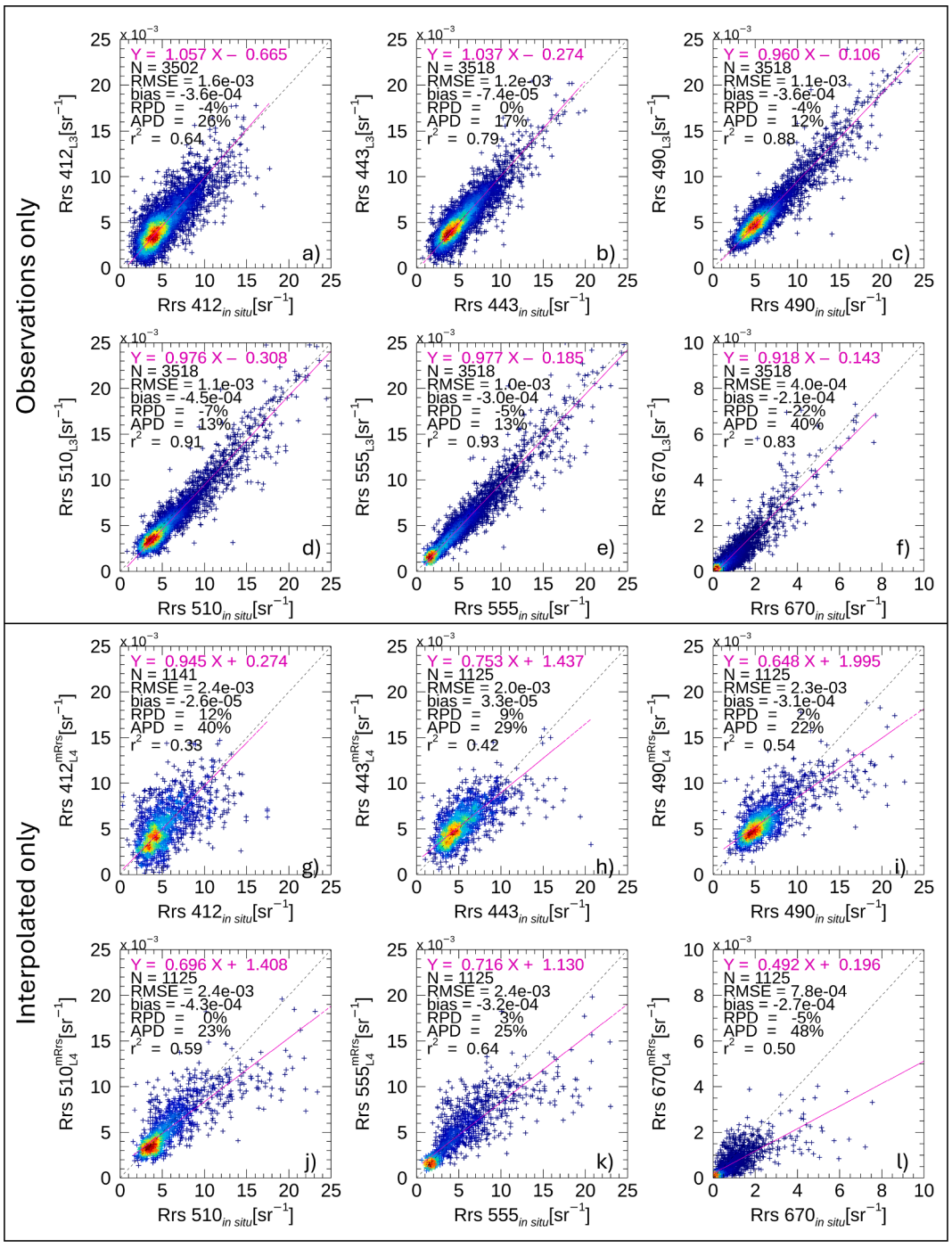


Fig. 7. Scatterplots of original L3 satellite Rrs observations against *in situ* Rrs measurements (panels a to f) and scatterplots of only interpolated pixels as derived from multi-band processing (L4) against *in situ* Rrs measurements (panels g to l).

showed a slope closest to 1 (0.975), indicating the best fit among the three plots (Fig. 8). As expected, these values tend to slightly worsen as interpolated pixels are grouped with observations (Fig. 8b) or considered alone (Fig. 8c) in the analysis; this is true for the regression line and RMSE and bias, although the correlation coefficient shows minor improvements, probably linked to the higher number of observations involved in two latter cases (i.e., Fig. 8b, 8c). The opposite is true when considering the relative comparisons between satellite and *in situ* data, with both APD and RPD improving from 21 % and 55 % in correspondence with observations (Fig. 8a) to 11 % and 48 % when observations and interpolated pixels are considered together (Fig. 8b) and to 4 % and 44 % when interpolated pixels considered alone (Fig. 8c). This inconsistency suggested a closer look at the relative percent difference distribution of the three plots of Fig. 8. The outcome is that these unexpected APD and RPD values are because the upper part of the data plots in Fig. 8 all show a higher degree of divergence, at least as compared to the lower parts. This along with the non-Gaussian distribution of these three data subsets translates in the average being not the best predictor for the comparison performance. Values in Table 4 show the percentiles of the relative difference distribution associated with the three plots of Fig. 8. The increase of the median relative difference (e.g., middle column of Table 4) with the inclusion in the analysis of the interpolated values (e.g., going from Fig. 8a to Fig. 8c) somehow meets the expectations, gives consistency back to the entire analysis and most

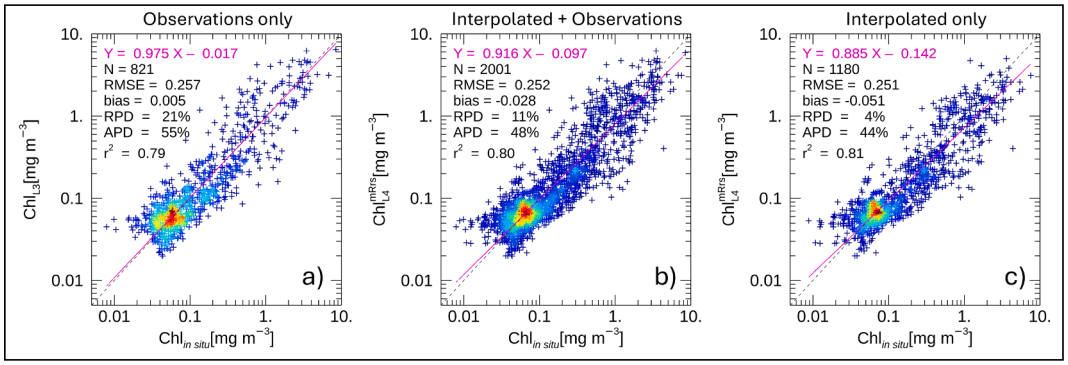


Fig. 8. Scatterplot of *in situ* Chl measurements against (a) *Chl*!" (Chl algorithm applied to original satellite *Rrs*!" observations), (b) to all available *Chl*!# *()'collocated data (Chl algorithm applied to original *Rrs*!" observations and available interpolated *Rrs*!# *()') and (c) to all available interpolated *Chl*!# *()'collocated data only (Chl algorithm applied only to interpolated *Rrs*!# '()') using the multi-band processing.

Table 4
Percentiles of the distribution of the relative difference (computed as the ratio between the $Chl_{Satellite}$-Chl_{insitu} difference and Chl_{insitu}) between the *in situ* and the satellite Chl data shown in Fig. 8. The second row refers to the comparison of *in situ* Chl against both Chl_{L3} (original observations) and Chl_{L4}^{mRrs} (interpolated values). For a complete description of the symbols, please refer to Table 1.

Satellite Data	Percentiles of relative % difference							
	0	5	25	50	75	95	100	
Chl_{L3}	−88.93	−55.41	−34.79	−0.31	47.05	169.08	776.54	Fig. 8a
Chl_{L3}&Chl_{L4}^{mRrs}	−90.74	−62.04	−35.63	−8.56	36.73	150.26	776.54	Fig. 8b
Chl_{L4}^{mRrs}	−90.74	−63.91	−36.37	−12.05	27.30	128.85	507.07	Fig. 8c

of all shows that the interpolation configuration developed here in this work is solid by providing excellent comparison against *in situ* observations.

Overall, these results show the effectiveness of the interpolation procedure and highlight that multi-band interpolation reduces the magnitude of percentage errors in the satellite-derived L4 Chl dataset, making it more reliable for further applications. However, the potential applications of DINEOF go beyond the specific dataset on Chl. Indeed, as we interpolated multispectral Rrs, similar improvements in accuracy could also be expected for other biogeochemical variables, such as a_{CDOM} and b_{bp}.

4. Summary and conclusions

Satellite remote sensing is valuable for monitoring OC parameters over extensive spatial and temporal scales. However, the amount of missing data in OC observations can be significant. Approximately 75 % of the ocean surface is usually cloud-covered, blocking visible electromagnetic radiation (Barth et al., 2024). This poses a challenge for satellite sensors relying on the visible light spectrum to measure OC data. In addition to cloud cover, data can be missing due to other issues such as atmospheric dust, sun glint contamination, and high sensor-zenith angles (Barth et al., 2024). The scarcity of biogeochemical data at various scales, from regional to global, underscores the need for broader coverage and improved detection of temporal and spatial variations. This issue extends beyond specific regions and encompasses various biogeochemical parameters, crucial for monitoring ocean ecosystem health and productivity and assessing the impacts of climate variability and change. Therefore, the effort made in this work was to evaluate an "upstream interpolation" approach for reducing data gaps in OC data time series.

In this context, the DINEOF technique has been explicitly assessed for the first time to reconstruct Rrs, the core parameter in OC remote sensing. Specifically, the novelty of this work stands in having interpolated the six Rrs_{L3} bands (i.e., 412, 443, 490, 510, 555, and 670 nm), following a multivariate-based approach. Employing DINEOF significantly enhanced data coverage of the daily Rrs dataset. As demonstrated by our results, the resulting gap-free Rrs dataset can be used to improve the accuracy of L4 Chl prediction. Furthermore, instead of performing EOF decomposition multiple times, for example, once for each Rrs band, the DINEOF multivariate approach does it all at once. This reduces the computational burden by avoiding redundant calculations and allows the algorithm to minimize processing time, making the multivariate approach more effective and efficient.

Moreover, we expect the accuracy of the Rrs_{L4}^{mRrs} to be further enhanced if the DINEOF algorithm exploits richer spectral information, mainly using hyperspectral reflectance data from, for example, new satellite missions such as PACE. Transitioning from multi-spectral to hyperspectral satellite sensors for ocean colour is expected, for example, to significantly improve satellite identification of phytoplankton biodiversity, a crucial component of the functioning of marine ecosystems and the upper ocean biogeochemistry. Hence, given the importance of using these data streams, this study underscores the utility and critical role of Rrs as a standalone dataset on which the DINEOF interpolation method can be applied to derive several gap-free biogeochemical parameters. From an operational point of view, the results of this paper demonstrate that using the DINEOF method with the proposed data processing configuration does not add any significant source of uncertainty and makes it possible to generate daily OC products routinely without any gaps. This may provide valuable global OC data, enhancing reliability and stability in operational satellite data processing workflow.

This study specifically targeted the Mediterranean Sea due to the large *in situ* dataset available for validation and its relatively low cloud coverage, which permitted well-designed experiments. Although the DINEOF multi-band interpolation method theoretically applies to any region, including coastal waters, tropical areas, subpolar regions, and even globally, its performance must be deeply evaluated (Zhao et al., 2024; Hilborn and Costa, 2018). As this work thoroughly highlighted,

the performance of any interpolation approach strongly depends on observation quality, distribution, and availability.

CRediT authorship contribution statement

Christian Marchese: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Simone Colella:** Writing – review & editing, Funding acquisition, Data curation. **Vittorio Ernesto Brando:** Writing – review & editing, Funding acquisition. **Maria Laura Zoffoli:** Writing – review & editing. **Gianluca Volpe:** Writing – review & editing, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been carried out within the framework of the Ocean Colour Thematic Assembly Centre of the Copernicus Marine Environment and Monitoring Service (contract: 21001L02-COP-TAC OC-2200 – Lot 2: Provision of Ocean Colour Observation Products (OC-TAC). The AERONET Team is acknowledged for their continuous support of the AERONET-OC subnetwork. Special thanks to Marco Talone from the Institut de Ciències del Mar (CSIC) and Giuseppe Zibordi, Barbara Bulgarelli, and Frederic Melin from the Joint Research Center of the European Commission for their role in establishing and maintaining the Venice, Acqua Alta, and Casablanca AERONET-OC sites. We also thank Vega Forneris and Flavio La Padula for their efforts in keeping the satellite data processing and archive at CNR-ISMAR. Two anonymous reviewers are warmly acknowledged for their comments. CM was an EMBL (Rome, Italy) internally funded postdoc and associate researcher at the CNR-ISMAR.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2024.104270.

Data availability

Data will be made available on request.

References

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J.M., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. Ocean Model. 9, 325–346. https://doi.org/10.1016/j.ocemod.2004.08.001.

Alvera-Azcárate, A., Barth, A., Beckers, J.-M., Weisberg, R.H., 2007. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. J. Geophys. Res. 112. https://doi.org/10.1029/2006JC003660.

Alvera-Azcárate, A., Barth, A., Sirjacobs, D., Lenartz, F., Beckers, J.M., 2011. Data Interpolating Empirical Orthogonal Functions (DINEOF): a tool for geophysical data analyses. Medit. Mar. Sci. 12, 5. https://doi.org/10.12681/mms.64.

Alvera-Azcárate, A., Barth, A., Parard, G., Beckers, J.-M., 2016. Analysis of SMOS sea surface salinity data using DINEOF. Remote Sens. Environ. 180, 137–145. https://doi.org/10.1016/j.rse.2016.02.044.

Alvera-Azcárate, A., Van Der Zande, D., Barth, A., Dille, A., Massant, J., Beckers, J.-M., 2024. Generation of super-resolution gap-free ocean colour satellite products using DINEOF. doi: 10.5194/egusphere-2024-1268.

Alvera-Azcárate, A., Vanhellemont, Q., Ruddick, K., Barth, A., Beckers, J.-M., 2015. Analysis of high frequency geostationary ocean colour data using DINEOF. Estuar. Coast. Shelf Sci. 159, 28–36. https://doi.org/10.1016/j.ecss.2015.03.026.

Alvera-Azcárate, A., Van der Zande, D., Barth, A., Troupin, C., Martin, S., Beckers, J.-M., 2021. Analysis of 23 Years of Daily Cloud-Free Chlorophyll and Suspended Particulate Matter in the Greater North Sea. Front. Mar. Sci. 8, 707632. https://doi.org/10.3389/fmars.2021.707632.

Arteaga, L.A., Rousseaux, C.S., 2023. Impact of Pacific Ocean heatwaves on phytoplankton community composition. Commun. Biol. 6, 263. https://doi.org/10.1038/s42003-023-04645-0.

Barth, A., Brajard, J., Alvera-Azcárate, A., Mohamed, B., Troupin, C., Beckers, J.-M., 2024. Ensemble reconstruction of missing satellite data using a denoising diffusion model: application to chlorophyll a concentration in the Black Sea. doi: 10.5194/egusphere-2024-1075.

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

Bélanger, S., Babin, M., Tremblay, J.-É., 2013. Increasing cloudiness in Arctic damps the increase in phytoplankton primary production due to sea ice receding. Biogeosciences 10, 4087–4101. https://doi.org/10.5194/bg-10-4087-2013.

Berthon, J.-F., Zibordi, G., 2004. Bio-optical relationships for the northern Adriatic Sea. Int. J. Remote Sens. 25, 1527–1532. https://doi.org/10.1080/01431160310001592544.

Blondeau-Patissier, D., Gower, J.F.R., Dekker, A.G., Phinn, S.R., Brando, V.E., 2014. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Prog. Oceanogr. 123, 123–144. https://doi.org/10.1016/j.pocean.2013.12.008.

Ćatipović, L., Matić, F., Kalinić, H., 2023. Reconstruction methods in oceanographic satellite data observation—a survey. JMSE 11, 340. https://doi.org/10.3390/jmse11020340.

Cazzaniga, I., Mélin, F., 2024. How Representative Are European AERONET-OC Sites of European Marine Waters? Remote Sens. (Basel) 16, 1793. https://doi.org/10.3390/rs16101793.

Cole, H., Henson, S., Martin, A., Yool, A., 2012. Mind the gap: the impact of missing data on the calculation of phytoplankton phenology metrics. J. Geophys. Res. Oceans 117, 1–8. https://doi.org/10.1029/2012JC008249.

Colella, S., Brando, V. E., Di Cicco, A., D'Alimonte, D., Forneris, V., Bracaglia M., 2023. Quality information document for Ocean Colour Mediterranean and Black Sea Observation Product. doi: 10.48670/moi-00299.

Concha, J.A., Bracaglia, M., Brando, V.E., 2021. Assessing the influence of different validation protocols on Ocean Colour match-up analyses. Remote Sens. Environ. 259, 112415. https://doi.org/10.1016/j.rse.2021.112415.

Di Sarra, A., Bommarito, C., Anello, F., Di Iorio, T., Meloni, D., Monteleone, F., Pace, G., Piacentino, S., Sferlazzo, D., 2019. Assessing the quality of shortwave and longwave irradiance observations over the ocean: one year of high-time-resolution measurements at the Lampedusa Oceanographic Observatory. J. Atmos. Oceanic Technol. 36, 2383–2400. https://doi.org/10.1175/JTECH-D-19-0018.1.

Enriquez-Alonso, A., Sanchez-Lorenzo, A., Calbó, J., González, J.-A., Norris, J.R., 2016. Cloud cover climatologies in the Mediterranean obtained from satellites, surface observations, reanalyses, and CMIP5 simulations: validation and future scenarios. Clim. Dyn. 47, 249–269. https://doi.org/10.1007/s00382-015-2834-4.

Gilerson, A., Herrera-Estrella, E., Foster, R., Agagliate, J., Hu, C., Ibrahim, A., Franz, B., 2022. Determining the Primary Sources of Uncertainty in Retrieval of Marine Remote Sensing Reflectance From Satellite Ocean Color Sensors. Front. Remote Sens. 3, 857530. https://doi.org/10.3389/frsen.2022.857530.

Gregg, W.W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. J. Mar. Syst. 69, 205–225. https://doi.org/10.1016/j.jmarsys.2006.02.015.

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J., Chuprin, A., Ciavatta, S., Cipollini, P., Donlon, C., Franz, B., He, X., Hirata, T., Jackson, T., Kampel, M., Krasemann, H., Lavender, S., Pardo-Martinez, S., Mélin, F., Platt, T., Santoleri, R., Skakala, J., Schaeffer, B., Smith, M., Steinmetz, F., Valente, A., Wang, M., 2019. Satellite ocean colour: current status and future perspective. Front. Mar. Sci. 6, 485. https://doi.org/10.3389/fmars.2019.00485.

Hilborn, A., Costa, M., 2018. Applications of DINEOF to satellite-derived chlorophyll-a from a productive coastal region. Remote Sens. (Basel) 10, 1449. https://doi.org/10.3390/rs10091449.

Hobday, A., Hartog, J., 2014. Derived ocean features for dynamic ocean management. Oceanography 27, 134–145. https://doi.org/10.5670/oceanog.2014.92.

Hollmann, R., Merchant, C.J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M., Wagner, W., 2013. The ESA climate change initiative: satellite data records for essential climate variables. Bull. Am. Meteorol. Soc. 94, 1541–1552. https://doi.org/10.1175/BAMS-D-11-00254.1.

Huot, Y., Antoine, D., 2016. Remote sensing reflectance anomalies in the ocean. Remote Sens. Environ. 184, 101–111. https://doi.org/10.1016/j.rse.2016.06.002.

Huot, Y., Antoine, D., Daudon, C., 2019. Partitioning the Indian Ocean based on surface fields of physical and biological properties. Deep Sea Res. Part II 166, 75–89. https://doi.org/10.1016/j.dsr2.2019.04.002.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., Claustre, H., 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies? doi: 10.5194/bgd-4-707-2007.

Konik, M., Kowalewski, M., Bradtke, K., Darecki, M., 2019. The operational method of filling information gaps in satellite imagery using numerical models. Int. J. Appl. Earth Obs. Geoinf. 75, 68–82. https://doi.org/10.1016/j.jag.2018.09.002.

Li, Y., He, R., 2014. Spatial and temporal variability of SST and ocean color in the Gulf of Maine based on cloud-free SST and chlorophyll reconstructions in 2003–2012. Remote Sens. Environ. 144, 98–108. https://doi.org/10.1016/j.rse.2014.01.019.

Li, M., Organelli, E., Serva, F., Bellacicco, M., Landolfi, A., Pisano, A., Santoleri, R., 2024. Phytoplankton spring bloom inhibited by marine heatwaves in the North-Western

Mediterranean Sea. Geophys. Res. Lett. 51 (20). https://doi.org/10.1029/2024GL109141.
Liu, X., Wang, M., 2018. Gap Filling of Missing Data for VIIRS Global Ocean Color Products Using the DINEOF Method. IEEE Trans. Geosci. Remote Sensing 56, 4464–4476. https://doi.org/10.1109/TGRS.2018.2820423.
Liu, X., Wang, M., 2019. Filling the Gaps of Missing Data in the Merged VIIRS SNPP/NOAA-20 Ocean Color Product Using the DINEOF Method. Remote Sens. (Basel) 11, 178. https://doi.org/10.3390/rs11020178.
Liu, X., Wang, M., 2022. Global daily gap-free ocean color products from multi-satellite measurements. Int. J. Appl. Earth Obs. Geoinf. 108, 102714. https://doi.org/10.1016/j.jag.2022.102714.
Marchese, C., 2015. Biodiversity hotspots: A shortcut for a more complicated concept. Global Ecol. Conserv. 3, 297–309. https://doi.org/10.1016/j.gecco.2014.12.008.
Marchese, C., Albouy, C., Tremblay, J.-É., Dumont, D., D'Ortenzio, F., Vissault, S., Bélanger, S., 2017. Changes in phytoplankton bloom phenology over the North Water (NOW) polynya: a response to changing environmental conditions. Polar Biol. 40, 1721–1737. https://doi.org/10.1007/s00300-017-2095-2.
Marchese, C., Castro de la Guardia, L., Myers, P.G., Bélanger, S., 2019. Regional differences and inter-annual variability in the timing of surface phytoplankton blooms in the Labrador Sea. Ecol. Ind. 96, 81–90. https://doi.org/10.1016/j.ecolind.2018.08.053.
Marchese, C., Hunt, B.P.V., Giannini, F., Ehrler, M., Costa, M., 2022. Bioregionalization of the coastal and open oceans of British Columbia and Southeast Alaska based on Sentinel-3A satellite-derived phytoplankton seasonality. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.968470.
Mayot, N., Matrai, P.A., Arjona, A., Bélanger, S., Marchese, C., Jaegler, T., Ardyna, M., Steele, M., 2020. Springtime Export of Arctic Sea Ice Influences Phytoplankton Production in the Greenland Sea. J. Geophys. Res. Oceans 125. https://doi.org/10.1029/2019JC015799.
McGinty, N., Guðmundsson, K., Ágústsdóttir, K., Marteinsdóttir, G., 2016. Environmental and climactic effects of chlorophyll-a variability around Iceland using reconstructed satellite data fields. J. Mar. Syst. 163, 31–42. https://doi.org/10.1016/j.jmarsys.2016.06.005.
Morel, A., Voss, K.J., Gentili, B., 1995. Bidirectional reflectance of oceanic waters: a comparison of modeled and measured upward radiance fields. *J. Geophys. Res. Oceans* 100 (C7), 13143–13150. https://doi.org/10.1029/95JC00531.
Navarro, G., Caballero, I., Prieto, L., Vázquez, A., Flecha, S., Huertas, I.E., Ruiz, J., 2012. Seasonal-to-interannual variability of chlorophyll-a bloom timing associated with physical forcing in the Gulf of Cádiz. Adv. Space Res. 50 (8), 1164–1172. https://doi.org/10.1016/j.asr.2011.11.034.
O'Reilly, J.E., Maritorena, S., Siegel, D.A., O'Brien, M.C., Toole, D., Mitchell, B.G., Kahru, M., Chavez, F.P., Strutton, P., Cota, G.F., Hooker, S.B., McClain, C.R., Carder, K.L., Muller-Karger, F., Harding, L., Magnuson, A., Phinney, D., Moore, G.F., Aiken, J., Arrigo, K.R., Letelier, R., Culver, M., 2000. Ocean Color Chlorophyll a Algorithms for SeaWiFS, OC2 and OC4: Version 4. In: Hooker, S.B., Firestone, E.R. (Eds.), SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. NASA Tech. Memo. 2000-206892 11. NASA Goddard Space Flight Center, Greenbelt, pp. 9–23.
Palacios, D.M., Bograd, S.J., Foley, D.G., Schwing, F.B., 2006. Oceanographic characteristics of biological hot spots in the North Pacific: a remote sensing perspective. Deep Sea Res. Part II 53, 250–269. https://doi.org/10.1016/j.dsr2.2006.03.004.
Polovina, J., Howell, E., 2005. Ecosystem indicators derived from satellite remotely sensed oceanographic data for the North Pacific. ICES J. Mar. Sci. 62, 319–327. https://doi.org/10.1016/j.icesjms.2004.07.031.
Racault, M.-F., Platt, T., Sathyendranath, S., A irba, E., Martinez Vicente, V., Brewin, R., 2014. Plankton indicators and ocean observing systems: support to the marine ecosystem state assessment. J. Plankton Res. 36, 621–629. doi: 10.1093/plankt/fbu016.
Rinaldi, E., Buongiorno Nardelli, B., Volpe, G., Santoleri, R., 2014. Chlorophyll distribution and variability in the Sicily Channel (Mediterranean Sea) as seen by remote sensing data. Cont. Shelf Res. 77, 61–68. https://doi.org/10.1016/j.csr.2014.01.010.
Taylor, M.H., Losch, M., Wenzel, M., Schröter, J., 2013. On the sensitivity of field reconstruction and prediction using empirical orthogonal functions derived from Gappy Data. J. Clim. 26, 9194–9205. https://doi.org/10.1175/JCLI-D-13-00089.1.
Volpe, G., Colella, S., Brando, V.E., Forneris, V., La Padula, F., Di Cicco, A., Sammartino, M., Bracaglia, M., Artuso, F., Santoleri, R., 2019. Mediterranean ocean colour Level 3 operational multi-sensor processing. Ocean Sci. 15, 127–146. https://doi.org/10.5194/os-15-127-2019.
Volpe, G., Buongiorno Nardelli, B., Colella, S., Pisano, A., Santoleri, R., 2018. An Operational Interpolated Ocean Colour Product in the Mediterranean Sea, in: Chassignet, E.P., Pascual, A., Tintoré, J., Verron, J. (Eds.), New Frontiers in Operational Oceanography. GODAE OceanView. doi: 10.17125/gov2018.ch09.
Volpe, G., Santoleri, R., Vellucci, V., Ribera d'Alcalà, M., Marullo, S., D'Ortenzio, F., 2007. The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. Remote Sens. Environ. 107, 625–638. https://doi.org/10.1016/j.rse.2006.10.017.
Volpe, G., Nardelli, B.B., Cipollini, P., Santoleri, R., Robinson, I.S., 2012. Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations. Remote Sens. Environ. 117, 223–235. https://doi.org/10.1016/j.rse.2011.09.020.
Waite, J.N., Mueter, F.J., 2013. Spatial and temporal variability of chlorophyll-a concentrations in the coastal Gulf of Alaska, 1998–2011, using cloud-free reconstructions of SeaWiFS and MODIS-Aqua data. Prog. Oceanogr. 116, 179–192. https://doi.org/10.1016/j.pocean.2013.07.006.
Yoder, J., Doney, S., Siegel, D., Wilson, C., 2010. Study of marine ecosystems and biogeochemistry now and in the future: examples of the unique contributions from space. Oceanog 23, 104–117. https://doi.org/10.5670/oceanog.2010.09.
Zhao, H., Matsuoka, A., Manizza, M., Winter, A., 2024. DINEOF interpolation of global ocean color data: error analysis and masking. J. Atmos. Oceanic Technol. 41 (10), 953–968. https://doi.org/10.1175/JTECH-D-23-0105.1.
Zibordi, G., Holben, B.N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D.M., Sorokin, M.G., 2021. Advances in the Ocean Color Component of the Aerosol Robotic Network (AERONET-OC). J. Atmos. Oceanic Technol. 38, 725–746. https://doi.org/10.1175/JTECH-D-20-0085.1.