Review

Reconstruction Methods in Oceanographic Satellite Data Observation—A Survey

Leon Ćatipović [1], Frano Matić [2] and Hrvoje Kalinić [1,*]

1 Environmental Data Analysis Laboratory, Faculty of Science, University of Split, 21000 Split, Croatia
2 University Department of Marine Studies, University of Split, 21000 Split, Croatia
* Correspondence: hrvoje.kalinic@pmfst.hr

Abstract: Oceanographic parameters, such as sea surface temperature, surface chlorophyll-a concentration, sea surface ice concentration, sea surface height, etc., are listed as Essential Climate Variables. Therefore, there is a crucial need for persistent and accurate measurements on a global scale. While in situ methods tend to be accurate and continuous, these qualities are difficult to scale spatially, leaving a significant portion of Earth's oceans and seas unmonitored. To tackle this, various remote sensing techniques have been developed. One of the more prominent ways to measure the aforementioned parameters is via satellite spacecraft-mounted remote sensors. This way, spatial coverage is considerably increased while retaining significant accuracy and resolution. Unfortunately, due to the nature of electromagnetic signals, the atmosphere itself and its content (such as clouds, rain, etc.) frequently obstruct the signals, preventing the satellite-mounted sensors from measuring, resulting in gaps—missing data—in satellite recordings. One way to deal with these gaps is via various reconstruction methods developed through the past two decades. However, there seems to be a lack of review papers on reconstruction methods for satellite-derived oceanographic variables. To rectify the lack, this paper surveyed more than 130 articles dealing with the issue of data reconstruction. Articles were chosen according to two criteria: (a) the article has to feature satellite-derived oceanographic data (b) gaps in satellite data have to be reconstructed. As an additional result of the survey, a novel categorising system based on the type of input data and the usage of time series in reconstruction efforts is proposed.

Keywords: data reconstruction; gap filling; missing data; gaps; satellite oceanography

1. Introduction

Oceanographic variables represent a significant subsection of Essential Climate Variables (ECV) [1–3]. Similar to all other ECVs, if oceanographic variables are to be used for analysis and forecasting purposes, persistent measuring and monitoring is required, both locally and globally. Generally, measuring methods can be divided into two distinct categories: in situ methods and *remote sensing* methods [4]. Depending on the specific measurement needs, one or the other method provides a better fit. For example, in situ methods, despite being difficult to perform accurately, are vastly superior compared to remote sensing methods when dealing with localised measurements, allowing for maximum accuracy and significantly higher spatial and temporal resolution [4]. However, scaling the in situ measurement coverage both spatially and temporally, in most cases, is exponentially difficult in regard to financial costs and logistical planning. This issue becomes even more prominent when one's aim is to achieve global coverage. This issue can be rectified by implementing remote sensing methods. Remote sensing encompasses measuring techniques based on analysing the electromagnetic signal retrieved from a target location [5]. Remote sensors are particularly useful when mounted on satellites [4,6], allowing for vast spatial coverage with relatively high temporal resolution.

As mentioned, the crux of remote sensing is the analysis of the detected electromagnetic signal. Depending on the target variable, a certain range of the electromagnetic spectrum is considered [4,6]. Furthermore, depending on the mentioned range, the measurement is carried out either actively or passively. Active measurement implies emission of the signal and detection of the subsequent reflected signal. Examples of active measurement include sea surface height and sea roughness. When measuring the two, a signal from the microwave range of the spectrum is emitted. The travel time and strength of the backscattered are measured and used as a starting point for the derivation of the aforementioned variables [4,6].

Not all oceanographic variables can be detected using the active measurement. Some of them rely on passive measurements: meaning the detection of naturally emitted and/or reflected signals (sunlight) from the sea. This infers that different parts of the electromagnetic spectrum are considered: visible light for ocean colour (and subsequently the concentration of chlorophyll-a suspended particulate matter, organic and inorganic carbon, coloured dissolved matter, etc.) [7,8] and infrared for sea surface temperature [9,10]. This, however, poses a problem, especially when considering satellite-based remote sensing because not all parts of the electromagnetic spectrum contain the same penetrative properties [11]. While the Earth's atmosphere is generally transparent for some wavelengths (such as microwaves), other parts of the spectrum can be significantly attenuated [11]. Due to non-penetrative nature of visible and infrared light cloud cover significantly diminished the area from which signals can be retrieved [12–15]. Other factors, such as unfavourable sun glint and high aerosol loading [16–18], further diminish spatial coverage. Even successful measurements are subject to removal in efforts of quality control [19–23]. All these negative influences result in gaps in satellite data, which hinders its usefulness.

Reducing the occurrences of gaps in satellite data are the next natural step. While improvements to the sensors [24,25] might reduce these gaps in future measurements, it will not help remove the gaps in already obtained data. Therefore, in the past two decades, significant efforts have been poured into various data-driven reconstruction techniques. By analysing existing available data, these techniques aim to infer the values of missing data in an effort to create complete, gap-free datasets. This paper provides a review of these methods when applied to oceanographic data. Among the data, sea surface temperature (SST) and surface chlorophyll-a concentration (chl_a) are significantly more prominent than the rest. Because some reconstruction approaches may simultaneously utilise data at various processing stages, next, Section 2 introduces the handling and derivation of satellite data, from raw data to final product. Section 3 briefly summarises the most popular gap-filling techniques used in satellite oceanography. Finally, Section 4 showcases examples of those techniques being applied in literature throughout the years.

2. Satellite Data

This section will briefly guide the reader through the nuances of what satellite data actually represents, how it is gathered and processed—from raw electromagnetic signals to mapped geophysical variables. In order not to digress too far from the main idea of the paper, only chl_a and SST products will be explored and explained.

2.1. Levels of Data

Obtaining measurements via satellites is an expensive and complex process—from the initial measurements, through signal conversion to the final data acquisition. While the exact nuances of this process may differ from one organisation to the other [26,27], the backbone procedure is mostly the same. Specifically, this subsection will shed light on the process implemented by NASA [26]. Depending on the process steps applied to the satellite data, the data are separated into specific stages called *levels*.

The first stage, level 0, refers to the initial content and format of data retrieved from the satellite after downloading the files to data cloud; this includes anything and everything that the satellite generates. These data have not been altered in any way, shape or form.

At this level, data are not usable for any scientific purposes, but the data are always saved as it can be used for recreating any subsequent data level in case of data loss. Additionally, the data are stored in specific packets, referred to as the Consultative Committee for Space Data Systems (CCSDS) packets [28].

Level 1 is usually subdivided into two sub-levels. Level 1A combines the CCSDS packets and restructures the data in units reflecting the way the sensor measures: for example, if the sensor in question is a scanner, then the unit of storage/structure is the set of pixels detected during that scan. Apart from measurement data, satellite navigation and other telemetry data are acquired. The final step of level 1A separates data into timed chunks (e.g., 5-min chunks), as a means of data size management [28]. Level 1B calibrates and geolocates level 1A data. Level 1A data comes in the form of integer counts. Calibration converts these counts into physical units, such as radiance/reflectance and brightness temperature. During calibration, instrument corrections are applied—nonlinear signal responses, temperature effects, sensor degradation and other instrument effects are all taken into account when considering instrument correction [4]. After the correction has been applied, navigation data of the spacecraft are combined with the geometry model of the instrument to determine the origin location of each measurement. This way, the data are geolocated [28]. Generally speaking, level 1 data are not an overly popular target for gap-filling [29–32].

Up to this point, satellite data can be viewed as "engineering data"—information on the spacecraft position and viewing angle, effects of instrument imperfections on data and so on. Level 2 data are the first instance of data being formatted into "scientific data". Level 2 data are the output of various algorithms, some of which will be explored in Sections 2.2 and 2.3, which take level 1B data as input. These algorithms may require ancillary data—data regarding different parameters, such as atmospheric water vapour, surface pressure, ozone layer thickness, etc., which are retrieved from sources outside the satellite in question itself. Level 2 data also contain quality flags—various parameters put in place to ensure data are within uncertainty standards set by the scientific community [28]. In the context of gap-filling, level 2 data are utilised for local/regional instances [33–57].

After level 2 data have been obtained, all instances of it from a single day can be combined spatially to create composites, creating level 3 data. While these composites usually contain a day's worth of information, composition can be carried out along the time axis, combining progressively longer time periods (8 days, month, season, year and finally entire duration of the satellite mission) [28]. Naturally, since level 3 data do not have spatial constrains, unlike level 2 data, it should be obvious that it is the most popular choice when dealing with data reconstruction [58–158].

The final step in data processing is the optional level 4. Level 4 data are optional in a sense that level 3 data are already well-suited and are used as-is by the scientific community. This level represents either a new variable type of data that has been derived from previously obtained satellite data (e.g., ocean primary productivity data are derived from chl_a data [26]) or satellite data that have been augmented in some way [26]. This augmentation can be achieved via an increase in spatial resolution [155,159–161], or, more popularly as will be showcased in Section 4, via missing data reconstruction. Some of the reviewed articles utilised level 4 data for gap-filling proof-of-concept purposes [162–164].

2.2. Surface Chlorophyll—A Concentration

Radiance and reflectance were mentioned in previous the subsection. In order to better grasp the physical characteristics of these variables, they are best explained through the design behind their respective sensors. chl_a is derived from ocean colour—visible light being reflected from the sea towards the satellite. Measuring the intensity of this light can be carried out via a Gershun tube radiometer [165]. The design of the radiometer is displayed in Figure 1. On one end of the tube is a hole of a fixed size which allows light to enter the device. Entering the hole, the light passes through the collection tube, which has internal light baffles. By building the hole and the collection tube to be of specific proportions,

along with the baffles, incoming light is filtered according to the incoming angle, denoted α. This angle is usually $5°$ to $10°$. If any light enters from an angle $\hat{\xi}$ greater than α, the light is blocked and does not affect the measurement. Knowing α, the viewing solid angle of the detector is given by $\Delta\Omega = 2\pi(1 - \cos\alpha)$ [165]. Passing through the collection tube, light is filtered according to wavelength λ. This filter allows for a certain bandwidth $\Delta\lambda$ to pass through. In most cases today, the bandwidth is around 10 nm, while hyperspectral instruments may shrink this to just a few nanometers [24,25,165]. Before hitting the photo-sensitive detector located at the end of the tube, light is diffused over an area ΔA. Knowing the properties of the light that is landing on the detector, radiance L is defined as [165]:

$$L(t, \hat{\xi}, \lambda) = \frac{\Delta Q}{\Delta t \Delta A \Delta\Omega \Delta\lambda} \quad \left(\text{W m}^{-2}\,\text{sr}^{-1}\,\text{nm}^{-1}\right). \tag{1}$$

Essentially radiance is the power of the light radiation at a specific wavelength per unit solid angle per unit area. Remove the collection tube dependence on the viewing angle and, instead of radiance, irradiance E is measured [165]:

$$E(t, \lambda) = \frac{\Delta Q}{\Delta t \Delta A \Delta\lambda} \quad \left(\text{W m}^{-2}\,\text{nm}^{-1}\right). \tag{2}$$

Depending onthe orientation of the device, the irradiance measured is usually either *downwelling* E_d or *upwelling* E_u [165].
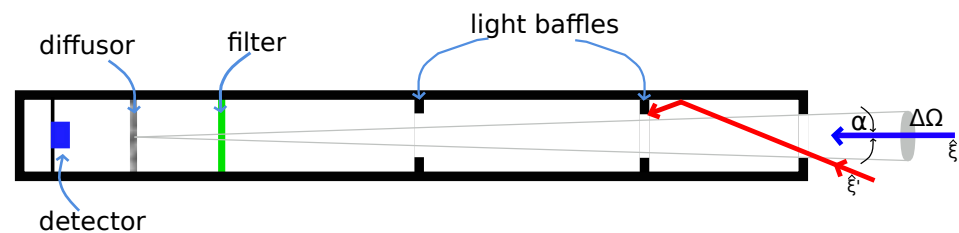

Figure 1. Cross-section of the Gershun tube radiometer.

Unfortunately, property of both radiance and irradiance is sensitive to receiving radiation due to sea surface perturbations and, as such, they usually are not implemented individually. However, since they exhibit a certain degree of correlation, their ratio attenuates this sensitivity enough to be usable for scientific purposes [165]. This ratio can be defined in multiple ways [165], but the most used one in ocean colour sciences is the so-called *remote-sensing reflectance* R_{rs} [165]:

$$R_{\text{rs}}(\theta, \phi, \lambda) = \frac{L_{\text{w,air}}(\theta, \phi, \lambda)}{E_{\text{d,air}}(\theta, \phi, \lambda)} \quad \left(\text{sr}^{-1}\right), \tag{3}$$

where the subscript air denotes evaluation just above the sea surface, and subscript w denotes water-leaving (or upwelling) radiance. Angles θ and ϕ denote the upward direction of the light exiting the water [165]. Combining this information with the level division provided in the previous subsection, it is prudent to note how raw radiance and raw irradiance are practically the only level 1 data available. Every variable derived henceforth, including water-leaving radiance L_{w} and normalised water-leaving radiance nL, is considered to be at least level 2 data as some algorithms (e.g., the atmospheric correction) need be applied [26,165].

With radiances and reflectances explained, chl$_a$ algorithms become much more comprehensible. Depending on the data provider, a variety of algorithms may be used, on both global and regional scales [156,166–174]. Most of these algorithms follow one of the two general formula shapes, depending on if they estimate chl$_a$ in oligotrophic [168] or eu-

trophic [166] waters. For example, NASA [26] offers a global weighted blend of both algorithms. The oligotrophic algorithm [168] is given as:

$$\text{CI} = R_{\text{rs}}(\lambda_{\text{green}}) - \left[R_{\text{rs}}(\lambda_{\text{blue}}) + \frac{\lambda_{\text{green}} - \lambda_{\text{blue}}}{\lambda_{\text{red}} - \lambda_{\text{blue}}} (R_{\text{rs}}(\lambda_{\text{red}}) - R_{\text{rs}}(\lambda_{\text{blue}})) \right], \tag{4}$$

where CI is the so-called *colour-index*, and λ_{red}, λ_{green} and λ_{blue} are the sensor appropriate wavelengths. CI is then used to estimate chl_{a}:

$$\log_{10}(\text{chl}_{\text{a}}) = a_0 + a_1\text{CI}. \tag{5}$$

For higher concentrations, the eutrophic (or the standard) algorithm is used [166]:

$$\log_{10}(\text{chl}_{\text{a}}) = a_0 + \sum_{i=1}^{4} a_i \left(\log_{10}\left(\frac{R_{\text{rs}}^{\text{MAX}}}{R_{\text{rs}}(\lambda_{\text{green}})} \right) \right)^i. \tag{6}$$

Here, $R_{\text{rs}}^{\text{MAX}}$ denotes the maximum value of R_{rs} between several different wavelengths [166]. Exact values of a_i depend on the scale and region where the algorithm is applied [166,170–172,174].

chl_{a} has been targeted for gap-filling via various methods since the early 2000s [84, 131,137], but the popularity soared in the past several years [30,31,33–46,50–53,57–83,85–93,132–136,138,148,155–157].

2.3. Sea Surface Temperature

Measurement devices developed for SST measurements are similar in design to the ones developed for chl_{a} [4,6,175]. The only difference between the two is, whereas the first measures visible light and derives radiance (and related variables), SST measuring devices measure other parts of the electromagnetic spectrum and derive brightness temperature B_λ. Measuring SST using remote sensing methods is based on Planck's law: every body emits a flux of energy, B_λ, proportional to the body's temperature, T, at a given wavelength λ. The flux is given by:

$$B_\lambda(\text{T}) = \frac{2hc^2\lambda^{-5}}{e^{hc/(\lambda k\text{T})-1}}, \tag{7}$$

where h is the Planck's constant, c is the speed of light, and k is Boltzmann's constant. Naturally, the spectrum of the flux is continuous regarding λ; however, this does not mean that measuring of sea's temperature can be achieved at any given wavelength. Generally, three conditions are considered when determining the target wavelength(s). Firstly, the sea should emit a detectable amount of energy given the usual temperature. Next, since the sensors are mounted on satellite flying above the atmosphere, the signal reaching the sensor should be as unaffected by the atmosphere as possible. Finally, measuring the flux at the given wavelength should be technologically possible, not to mention practical [4,175]. All things considered, SST can be measured in the infrared and microwave part of the spectrum [4,175,176]. Both wavelength choices have their advantages and disadvantages: while infrared measurements are more accurate and achieve higher resolution, microwave measurements are more robust in terms of measuring through cloud cover [177,178]. Generally, infrared measurements are more prominent [179–186]. However, the entire infrared spectrum not is used for SST measurement. Typically, infrared sensors are designed to measure between 3–5 μm and 8–12 μm [4,187]. These spectra are usually called channels [4]. Assuming sensors are well-calibrated, the next step in SST retrieval is removal of effects of the atmosphere and its contents have on the signal. Simplified, an assumption can be made that the difference between the energy fluxes detected between two wavelengths is related to the temperatures difference between them [4,175,187]:

$$\text{SST}_{ij} - \text{T}_i = f(\text{T}_i - \text{T}_j), \tag{8}$$

where subscripts i and j refer to the sensor channels. Considering that the atmospheric effects are small in these particular channels, SST can be estimated by a simple linear function [4,180,182,184,185,187,188]:

$$\text{SST}_{ij} = a_0 + a_i \text{T}_i + a_j \text{T}_j, \tag{9}$$

where a_x are coefficients derived by regression analysis or in situ comparisons [4,189]. Expanding upon this linear estimation, higher accuracy can be achieved if sensor viewing angle is considered [4,183,184,186,187]:

$$\text{SST}_{ij} = b_0 + b_1 \text{T}_i + b_2 (\text{T}_i - \text{T}_j) \text{SST}_r + b_3 (\text{T}_i - \text{T}_j)(\sec \theta - 1), \tag{10}$$

where b_x are the new coefficients, SST_r is the basic linear estimation of SST, and θ is the sensor zenith angle. Further improvements of the algorithms are mostly derived from the presented ideas. Depending on the specific sensors and product suite, different algorithms may be utilised. For example, NASA's ocean colour suite [190] offers a variety of SST products obtained by different algorithms [183,186].

In comparison to chl$_a$, SST received slightly less reconstruction attention [43,44,46–49, 52,54–56,87–91,94,105–128,140–147,150–154,162,191–194]. However, examining the timeline of the development of reconstruction methods, one can notice that, in some cases, SST is the first variable to be targeted with novel methods, sort of as a proof-of-concept. This should not necessarily be surprising, as SST is slightly more trivial to derive than chl$_a$, is significantly less volatile than chl$_a$, and microwave SST is significantly more impervious to atmospherically-induced gaps.

2.4. Other Satellite Derived Oceanographic Variables

While chl$_a$ and SST are the most popular oceanographic variables, other types of variables are also available and have also been targeted for reconstruction. These include, but are not limited to: sea surface height/sea level anomaly (SSH/SLA) [100–104,139,158] sea surface salinity (SSS) [29,33,54], total suspended matter/suspended particulate matter concentration (TSM/SPM) [45,52,130,148,195], photosynthetically active radiation (PAR) [86,87], diffuse attenuation coefficient (K_d) [30,40,148,149], phytoplankton and particulate concentration (a_{ph}, b_{bp}) [53,83,136] and so on [31,32,41,42,96,99,129,163,164].

3. Reconstruction Methods

Before delving into the details of various reconstruction methods, it is prudent to define several categories of reconstruction approaches to better make sense of similarities and dissimilarities between the approaches. Our findings show that it is possible and practical to separate the methods (and the subsequent application of the respective methods in literature) regarding two criteria: type of data used and time-instances of data used. Type of data used refers to whether or not reconstruction of target data has been achieved using additional, relevant, non-target type of data. The additional, non-target type data are referred to as *proxy* data. Proxy data may improve reconstruction accuracy if there is significant correlation between the target and proxy data. Mathematically, let x be the target data. Then, in a single time-instance of satellite data (a single satellite image), let r_a represent areas of the image containing available data and r_m areas of the image with missing data, so that:

$$r = r_a \oplus r_m \tag{11}$$

represents the entire area of the image, where $\oplus$ denotes the operation of spatial composition. Let f be some reconstruction function that, for a given input, outputs the value of the target data in missing areas. Finally, let y be a proxy data. Then, two approaches may be defined:

$$x(r_m) = f(x(r_a)), \tag{12}$$

and

$$x(r_{\mathrm{m}}) = f(x(r_{\mathrm{a}}), y(r_{\mathrm{m}} \lor r)) \tag{13}$$

or

$$x(r_{\mathrm{m}}) = f(y(r_{\mathrm{m}} \lor r)), \tag{14}$$

where $\lor$ denotes the logic operation "or". Equation (12) represents the non-proxy approach, while both Equations (13) and (14) represent the proxy approach. These approaches may also be called the univariate and multivariate approaches. An example of this distinction between the two is illustrated in Figure 2. For clarity, these examples are limited to one proxy data, but no such limit need exist in reality.
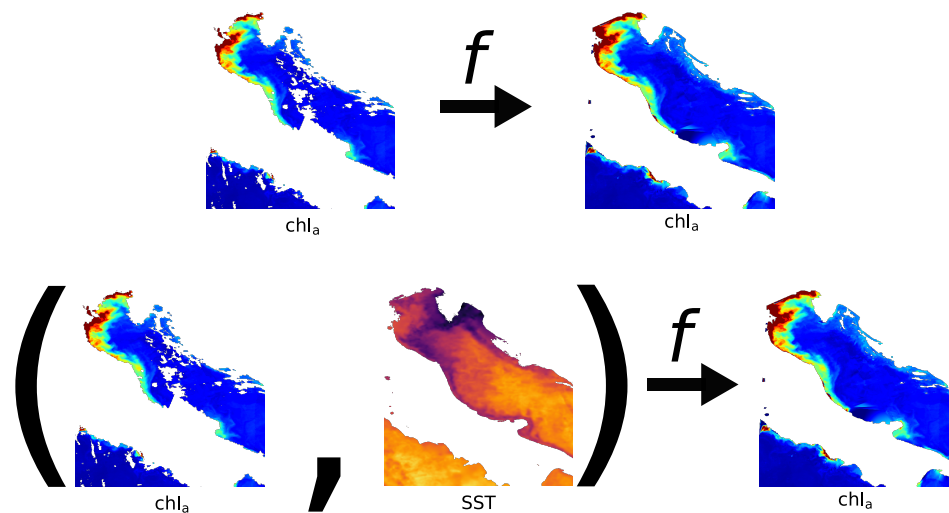
Figure 2. Examples of univariate and multivariate reconstruction approaches. In both cases, chl_{a} is the target data x. Missing data in this case were caused by rough sea surface. The top row exemplifies the non-proxy approach. The bottom row exemplifies the proxy approach, where SST has been taken as proxy data y. Both examples imply available data have been composited with reconstructed missing data to obtain complete data, $x(r) = x(r_{\mathrm{a}}) \oplus x(r_{\mathrm{m}})$.

Similarly, data can be reconstructed with or without drawing resources from time series. Let t_{n} denote the targeted image time-instance. The single time-instance approach is virtually identical to the univariate approach from Equation (12):

$$x(r_{\mathrm{m}}, t_{\mathrm{n}}) = f(x(r_{\mathrm{a}}, t_{\mathrm{n}})). \tag{15}$$

However, the approach with multiple time-instances is then given by:

$$\begin{aligned} x(r_{\mathrm{m}}, t_{\mathrm{n}}) = &f(x(r_{\mathrm{a}}, t_{\mathrm{n}-q_1}), x(r_{\mathrm{a}}, t_{\mathrm{n}-(q_1-1)}), \\ &\dots, x(r_{\mathrm{a}}, t_{\mathrm{n}}), \dots, x(r_{\mathrm{a}}, t_{\mathrm{n}+(q_2-1)}), x(r_{\mathrm{a}}, t_{\mathrm{n}+q_2})), \end{aligned} \tag{16}$$

where q_1 and q_2 denote the number of considered time-instances prior and after the targeted time-instance. In practical terms, since most satellite data have a temporal resolution of a single day, time-instances usually refer to day; if t_{n} is 15 July 2014, then $t_{\mathrm{n}-1}$ denotes 14 July 2014, $t_{\mathrm{n}+1}$ denotes 16 July 2014 and so forth. Distinction between the two latter approaches is depicted in Figure 3. If the reconstruction is completed using information from a time series, the reconstruction is simply referred to as multitemporal, but if it draws resources from only one time-instance, it is referred to as unitemporal.
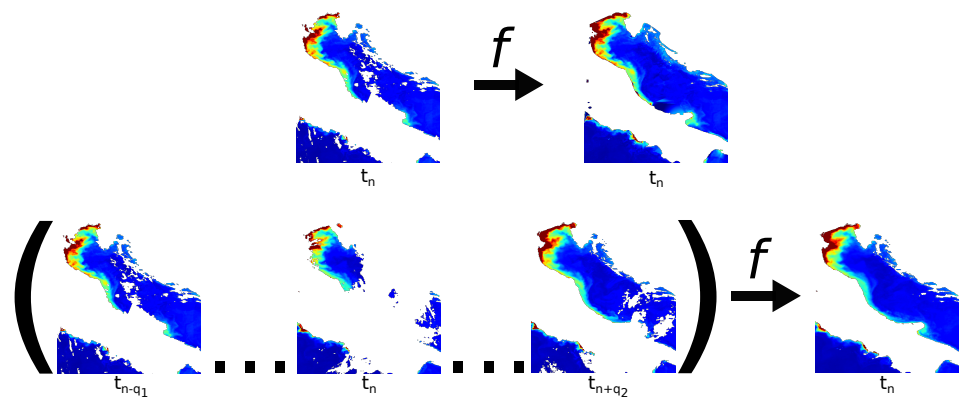
Figure 3. Examples of unitemporal (top row) and multitemporal (bottom row) approaches being used for chl$_a$ reconstruction. The targeted time-instance is t_n. Both examples imply available data have been composited with reconstructed missing data to obtain complete data, $x(r, t_n) = x(r_a, t_n) \oplus x(r_m, t_n)$.

It is important to note how the two criteria are not mutually exclusive: a method may be categorised as any one of the four possible combinations: e.g., simple spatial interpolation of a single chl$_a$ swath would be placed in the Univariate/Unitemporal category, while spatio-temporal interpolation of combined chl$_a$ and SST data would be played in the Multivariate/Multitemporal category. It is also important to point out that these categories are not always rigid. As will be seen in the following subsections, one method might be originally developed for univariate reconstruction and subsequently minimally modified to allow for multivariate reconstruction. Therefore, rather than putting emphasis on sorting the methods themselves into these categories, it is more sensible to focus on sorting the actual reconstruction approaches, meaning the combination of method and input data. To put this categorisation system in place, Table 1 has been added.

Table 1. Reconstruction approaches sorted according to the novel variate-temporal categorisation system.

	Unitemporal	Multitemporal
Univariate	[35,38,41,42,48,51,59–61,68,74–76,79,80,84,97–99,105,108,113–117,122,124,126,129,137,138,141,148,152,153,156,158,163,192]	[30–32,34,37,39,40,43–45,47,50–52,54,55,57–59,63,64,66,68,69,71–74,76–79,81–83,85–87,91,93,100,101,104,105,107–113,117,119–121,125,127,128,130,133,135,139,140,142–144,147,148,150–152,154,155,158,162–164,193,193–195]
Multivariate	[29,36,46,49,53,56,61,62,65,67,94,106,118,123,131,132,136,145]	[29,64,67,70,88–90,92,93,95,96,102,146,157,195]

The remainder of this section briefly presents the main idea behind the most popular reconstruction methods presented in literature. Naturally, explaining every detail and modification of every existing variation would detract too much from the main idea of this paper. Thus, should the reader wish to inquire more about a particular reconstruction method, they should refer directly to source material.

3.1. EOF/DINEOF

Empirical orthogonal functions (EOF) have been hypothesised to be useful in data interpolation [151]. EOF are calculated using singular value decomposition (SVD) [196]. A matrix X containing observations so that X_{ij} represents the value of the field $f(r, t)$ at location r_i at the moment t_j, i.e.:

$$X_{ij} = f(r_i, t_j), \tag{17}$$

can be written as a succession of n column vectors x_j size of m at moment t_j:

$$X = (x_1, x_2, \dots, x_n). \tag{18}$$

This matrix may represent a set of satellite images, among other types of data [151]. Then, applying SVD, the u and v eigenvectors (sizes of m and n) are obtained:

$$
\begin{aligned}
Xv &= \rho u \\
X^* u &= \rho v,
\end{aligned}
\tag{19}
$$

where ρ is the singular value, and X^* is the adjoint of X. The decomposition is equivalent to:

$$
\begin{aligned}
XX^* u &= \rho^2 u \\
X^* Xv &= \rho^2 v.
\end{aligned}
\tag{20}
$$

By decomposing this way, u and v can be thought of as the eigenvectors of time-averaged covariance matrix XX^* and the spatially-averaged covariance matrix X^*X [151]. Since the eigenvectors are normalised, initial matrix X can be decomposed into:

$$
X = UDV^* = \sum_{k=1}^{q} \rho_k u_k v_k^*,
\tag{21}
$$

where U and V are the matrices composed of columns of u_i and v_i, and D is the diagonal matrix so that $D_{ij} = \rho_i \Delta_{ij}$, where Δ_{ij} is the Kronecker symbol. SVD allows for accurate approximation of X by using just the first k dominant eigenvectors instead of the complete set of q eigenvectors [197]. SVD is only applicable on complete matrices—datasets that have no missing values. This, as stated before, is not a given in the domain of satellite oceanography. Since EOF aims to describe the signal, combinations of dominant EOF might be used to estimate missing values [151]. In order to apply EOF for this purpose, a certain setup is needed. Firstly, let I denote the set of missing data, containing n_0 missing data. X_t is the reconstructed dataset, X_o is the observed dataset with missing data, and X_a is the dataset with missing data filled in. X_o is obtained by setting the missing values (i, j) of I to 0. After that, SVD is applied to X_o, so that [151]:

$$
UDV^* = X_o.
\tag{22}
$$

Then, using the first N dominant eigenvectors, interpolated values of X_a can be obtained [151]:

$$
(X_a)_{ij} = (U_N D_N V_N^*)_{ij} = \sum_{k=1}^{N} \rho_k (u_k)_i (v_k^*)_j, \ (i, j) \in I.
\tag{23}
$$

U_N and V_N^* consist of N first spatial and temporal EOF. Now, X_a can be written as:

$$
X_a = X_0 + \Delta X,
\tag{24}
$$

where ΔX is zero for all non-missing data points. This procedure can be repeated again:

$$
UDV^* = X_a
$$
$$
(X_{a1}) = (U_N D_N V_N^*)_{i,j} = \sum_{k=1}^{N} \rho_k (u_k)_i (v_k^*)_j, \ (i, j) \in I,
\tag{25}
$$

where X_{a1} denotes filled data that have been subjected to mentioned estimation twice. Theoretically, this procedure can be repeated indefinitely. In practice, this procedure goes on until convergence of estimation error based on a cross-validation technique [193]. This method is more popularly known as Data INterpolating Empirical Orthogonal Functions (DINEOF) [193].

DINEOF, its variations and other EOF-based reconstruction methods are arguably the most popular reconstruction methods: its development sparked a boom in data reconstruction, and it has served in many instances as benchmark for reconstruction accuracy com-

parison [29–31,33,34,37,39,40,43–45,50–52,55,58,59,63,66,69–74,77–79,81,82,85,87–93,95,101,102,104,105,107,119,120,125,127,128,130,133,135,142–144,148,150,151,157,162,191,193–195].

3.2. Support Vector Regression

One of the older methods utilised for dealing with regression problems is Support Vector Regression (SVR) [198]. Support Vectors, as the name suggest, represent data points which serve as a basis for decisions implemented for the rest of the data: from classification to regression, etc. [198,199]. To explain this approach, let x_n be a set of independent variables obtained during N measurements and let y_n be a set of corresponding dependent (or target) variables. SVR then primarily aims to determine the function f which combines them. For a linear problem, this function is expressed as:

$$y_i = f(x_i), \ i \in 1, \ldots, N, \tag{26}$$

so that the error margin between the actual measurements and estimations is ϵ. Explicitly, let f be:

$$f(x) = a \odot x + b, \tag{27}$$

so that a, x and b represent vectors, and $\odot$ represents scalar product. The secondary aim of SVR is to minimise the slope a. This can be achieved through minimising the norm $\|a^2\| = a \odot a$, so that the optimal hyperplane between the data can be achieved [200]. With these two conditions in place, SVR is essentially an optimisation problem:

- min $\frac{1}{2}\|a^2\|$ s. t.;
- $|y_i - (a \odot x_i + b)| \leq \epsilon$.

The formulation assumes both conditions can be fulfilled simultaneously, which may or may not be possible for a given problem [198]. To rectify this, slack variables ξ_i and ξ_i^* are introduced, practically making the error margin ϵ flexible [201]. Visualisation of ϵ and ξ is presented in Figure 4. Optimisation problem now becomes:

- min $\frac{1}{2}\|a^2\| + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$ s. t.;
- $y_i - (a \odot x_i) - b \leq \epsilon + \xi_i$;
- $(a \odot x_i) + b - y_i \leq \epsilon + \xi_i^*$ and
- $\xi_i, \xi_i^* \geq 0$.

This is the backbone of SVR [201]. SVR can be applied to nonlinear problems; delinearisation is achieved via Lagrange multipliers [202] and by introducing kernels [198,203,204].
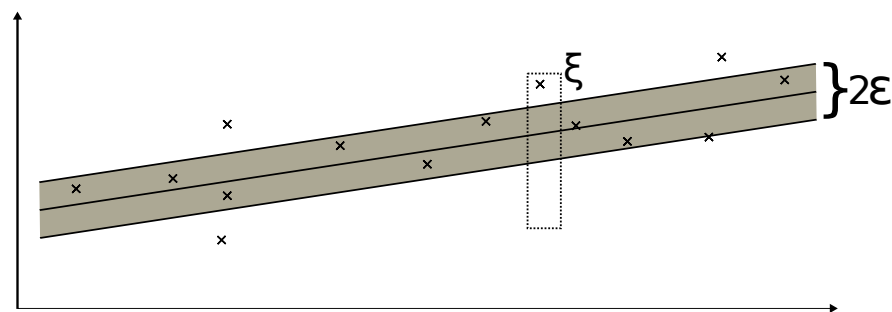
Figure 4. Error margin ϵ in regard to a linear problem. The dotted box represents the slack variable ξ, allowing data outside the grey error margin to influence the attributes of function f.

SVR has seldom been used in the context of satellite oceanographic data reconstruction [56,74,84], but it is worth noting in the context of machine-learning-based reconstructions.

3.3. Kriging/Optimal Interpolation

Let X represent a random geophysical variable and its realisation x_1 at a random location s_1. The space A contains N such realisations $X(s_1), X(s_2), \ldots, X(s_N)$. Furthermore, let $X(s_m)$ represent the value at a location within A that is not available, effectively a

missing data point. Even though these variables may be random, there is some underlying geophysical correlation between them [205–207]. At this point, kriging (also known as Optimal interpolation OI) functions if two assumptions can be made [205–207]: stationarity and that the correlation between the two variables depends only on the distance between them and not on their actual location. Stationarity implies that any two subspaces of A contain the same statistics (i.e., mean and standard deviation). If those two assumptions are reasonable, then kriging can be written as [205–207]:

$$X(s_m) = \sum_{i=1}^{N} w_i X(s_i), \tag{28}$$

where w_i is a weight corresponding to variable $X(s_i)$. To determine the w_i, a function called variogram (γ) is needed. One such variogram is depicted in Figure 5. Formally, variogram can be written as [205,206]:

$$\gamma(s_i, s_j) = \frac{1}{2}(x_i - x_j)^2. \tag{29}$$

Effectively, variogram indicates that the closer two points s_i and s_j are, the difference between their respective variables should be lower and vice versa. Using the variogram, weights w_i can be determined by solving the matrix equation [207]:

$$Tw = R, \tag{30}$$

or its transpose:

$$w = T^{-1}R, \tag{31}$$

where w is the vector comprised of w_i, T is the matrix comprised of $\gamma(s_i, s_j)$ and R is the matrix consisting of $\gamma(s_m, s_i)$ [207]. Naturally, the estimation may not only be carried out spatially but also temporally [205,206]. Interpolation-based methods are amongst the oldest methods to be utilised in gap-filling in satellite oceanography [100,109] and have remained somewhat popular throughout the years [38,47,54,57,60,64,67,73,76,83,86,99,108,110–113, 117,120,121,132,140,152–154,157,158,162–164,191,193].
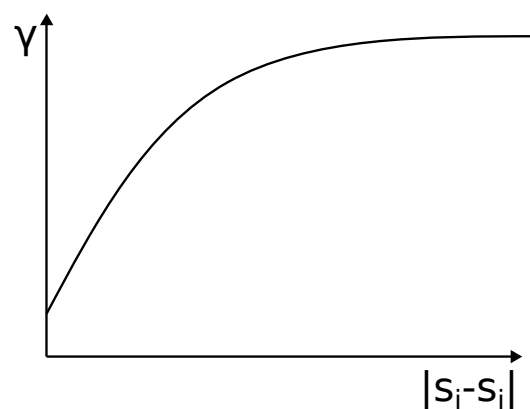
Figure 5. A theoretical variogram. The further two points are, their correlation becomes less and less significant up until a final cutoff value.

3.4. K-Nearest Neighbours

K-nearest neighbours (KNN) is a machine learning regression-solving method [208]. Regression takes form based on the sum of the K closest available data points. The sum can take many different forms, but it is most commonly expressed as a weighted function or as uniform [208]. In case of a weighted sum, increased distance equates to diminished influence. The distance is usually Euclidean, but if KNN is used in a geophysical domain, the distance may be the actual geographical distance [208]. Just like SVR, KNN is

not overly popular [67,74,163,164], but as a part of machine-learning reconstruction, it is also noteworthy.

3.5. Random Forest/Decision Tree

Random Forest (RF) stems from the Decision Tree (DT) algorithm. Although primarily used for classification problems, RF exhibits a capacity for regression problems as well [209]. Starting with DT, let $\bar{x}_i = (x_i^{(1)}), x_i^{(2)}, \ldots, x_i^{(m)}$ be a vector or a single point in data space. The vector components $x_i^{(j)}$ are then its features. DT functions by diving the data space into arbitrary classes according to various conditions based on the data features. With sufficiently enough conditions (so-called *nodes*), DT effectively divides the data space into enough classes so that each and every outcome of the nodes (so-called *leaf*) corresponds to an individual data point [209,210]. A simplified DT is depicted in Figure 6. RF is essentially made up of numerous DTs. Decisions made this way are the result of majority of decisions obtained from individual DTs. This improves stability and accuracy [209]. RF solves regression problems by first modelling a forest which accurately connects independent and target variables. Once the RF has been sufficiently trained, RF is provided with available independent variables and is tasked with determining the missing dependent variables [211].
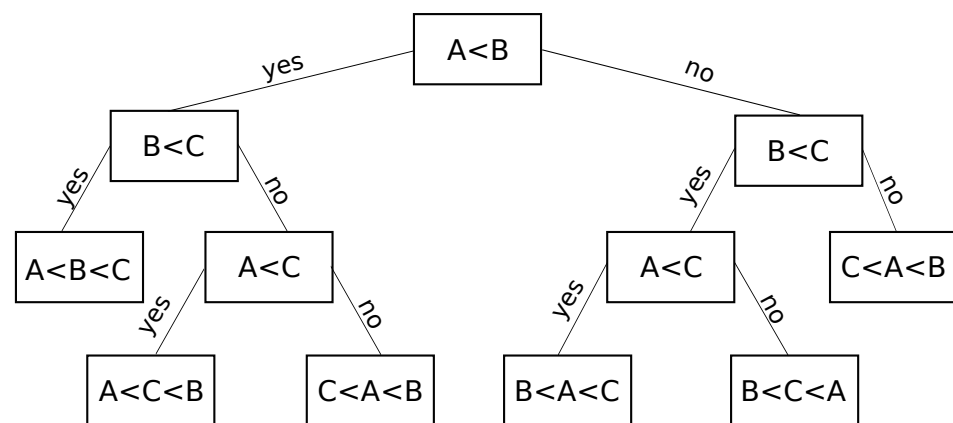
Figure 6. A simple DT for sorting the numbers A, B and C, assuming no two are equal. Since the DT is well defined, every leaf corresponds to possible sorting order.

RF-based (including DT) reconstructions have appeared consistently during the past several years [56,62,67,74,75,106,157,163].

3.6. Self-Organising Maps

Self-Organising Maps (SOM) is an unsupervised learning method typically used in classification problems [212]. However, it may be utilised for data reconstruction purposes [64]. SOM operates on neuron mesh, of a predetermined shape and structure [212]. The idea of SOM is as follows: let M denote a data manifold containing vectors v. This manifold can be described via the aforementioned mesh by moving the neurons (also known as Best Matching Units *BMU*), so that the distance between the individual *BMU*s and the vectors v $d(v, BMU_i) = |v - BMU_i|^2$ is minimal [212]. Depending on the mesh, there are N *BMU*s available. Should the data vectors v be distributed by the probability distribution $P(v)$ within the data manifold M, the reconstruction error L is given by:

$$L = \int P(v)d(v, BMU_i)dv. \tag{32}$$

Effectively, SOM functions by moving the *BMU*s through data space, reshaping the mesh to fit the manifold as closely as possible. However, since the *BMU*s are located on a "fixed" mesh, forming a strong correlation between individual *BMU*s, so that moving one *BMU*

affects the neighbouring *BMU*s [212]. The mesh often comes in a form of two-dimensional lattice [212]. Utilising this concept, SOM can be trained to describe typical oceanographic (or any other type of data) data manifolds by learning on complete datasets. Similarly to RF, once the model has been sufficiently trained, SOM can be used to predict the missing values based on available data [64,96,157], but unlike RF, SOM's popularity is more akin to the popularity of SVR and KNN.

3.7. Neural Networks

Neural networks (NN) are a subsection of machine learning methods with wide applicability. Traditionally, NN are introduced by examining the perceptron model [213]. Perceptron is a function which intakes input variable x and outputs the output variable $f(x)$. One of the simpler transformations is:

$$f(x) = \begin{cases} 1 & \text{if } a \odot x + b > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

where a is the weight, and b is bias [213]. After transformation, the output $f(x)$ is compared to the desired output value y. Depending on this comparison, the weight is updated, and the process repeats itself until the error between $f(x)$ and y is acceptable [213]. By combining these transformative functions in succession, deep neural networks are constructed [214,215]. The aforementioned functions now act as layers and can be grouped into the input layer, the output layer and the convoluted layer. The way the layers are ordered and organised is customisable, allowing for creation of various architectures of NN, such as Feed-forward NN [216] and Long Short-Term Memory [217]. Among these architectures are also the fairly new Generative Adversarial Networks (GAN) [218] and AutoEncoders (AE) [219].

3.7.1. Generative Adversarial Networks

GAN actually consists of two separate networks: a generating network, or the generator (G), and a discriminating network, or the discriminator (D). The aim of the generator is to generate a distribution p_g to be as similar as possible to the distribution p_x of real data x. For a given input z from distribution p_z, the generator generates $G(z)$ as output. Discriminator, on the other hand, has to determine whether $G(z)$ originated from p_x or p_z. This decision comes in the form of scalar $D(G(z))$. Therefore, the discriminator aims to become as accurate as possible when discriminating between "real" and "generated" data, while the generator aims to generate data as similar as possible to "real" data [218]. Mathematically, this is a *min-max* optimisation of the so-called *loss* function $V(G, D)$ [218]:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_x}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]. \tag{34}$$

3.7.2. AutoEncoders

AE function on the principle of information bottleneck [219,220]. AE consists of three layers: the encoder, the bottleneck and the decoder. The idea behind AE is based on dimensionality reduction: reducing the number of features required for data description. The encoder e is tasked with determining the reduced dimensionality, so that [219,220]:

$$\begin{aligned} x &\in R^n \\ x \to e(x), \ e(x) &\in R^m, \ \text{s.t.} (m < n). \end{aligned} \tag{35}$$

This way, point data x from n-dimensional data space R^n is transcribed into m-dimensional latent data space R^m. Effectively, the data are being encoded. Decoding of the data is achieved by the decoder d [219,220]:

$$\begin{aligned} e(x) &\in R^m \\ e(x) \to d(e(x)), \ d(e(x)) &\in R^n, \end{aligned} \tag{36}$$

which is nothing more than returning the data back into the original n-dimensional data space R^n. Finally, the bottleneck is tasked with further reducing the dimension m, so that the final dimension of the latent data space is lower than needed for direct reconstruction. AE aims to reduce the difference [219,220]:

$$loss = ||x - d(e(x))||^2 \tag{37}$$

utilising the noise variable ϵ:

$$loss = ||x - d(e(x + \epsilon))||^2. \tag{38}$$

In terms of data reconstruction, ϵ takes shape of missing data. AutoEncoder is the basis of a popular reconstruction method called DINCAE: Data INterpolating Convolutional AutoEncoder [94,105]. Similarly to DINEOF, DINCAE effectively reduces dimensionality in the input data, but whereas DINEOF relied on EOF decomposition, DINCAE relies on the convolutional layers of the encoder and the bottleneck [105].

NN-based reconstructions have seen a recent rise in popularity [49,56,59,61,65,70,74, 94,105,106,115,120–124,134,137,139,147,163,192].

### 3.8. Data Merge and Other Reconstruction Methods

Apart from the listed methods, there are several other methods useful for reconstruction. Out of these, the most prominent is data merging—merging data from various sources [35,41,42,46,48,51,68,79,97,98,108,114,117,131,148,149,152,158]. Technically, this approach is better referred to as gap-filling, but for the purposes of this paper, data merging was still taken into consideration. The remainder of these methods refers to various numerical models, filters, composites, spatial scaling, etc. [29,32,36,53,80,103,116,118,126,129,136, 138,141,145,155–157,192].

## 4. Summary Review of Reconstruction Methods in Literature

Presented in this section are more than 130 articles surveyed in preparation of this paper. During research, papers were selected by two criteria. The first criterion is that the paper has to feature any oceanographical satellite data. As mentioned earlier, these data mostly refer to chl$_a$ and SST, but other types of oceanographic data are also considered. The second criterion is that the gaps in the obtained satellite data have to be filled using any viable method. This criterion includes articles that propose novel gap-filling methods—i.e., development articles, but also any article that fills gaps in the initially retrieved data before further analysing and processing the data in ways that do not necessarily have to deal with the issue of missing data reconstruction—i.e., application articles. Naturally, out of all the surveyed articles, more are classified as application articles than as development articles. The statistics of survey covering the source journals are presented in Appendix A. Considering the number of articles reviewed, it is impractical to thoroughly display all nuances of each and every one of them. Therefore, articles were surveyed in the most concise manner possible, extracting parameters relevant to the majority of the articles. Due to the sheer size of the findings, the resulting Table A1 is placed in Appendix B.

To illustrate the popularity of reconstruction methods throughout the years from Table A1, Figure 7a,b have been added. Methods have been grouped as defined in Section 2. While EOF-based methods are the most popular ones, during the past several years, NN-based (and machine-learning-based in general) methods started to be utilised more consistently and gained certain popularity.
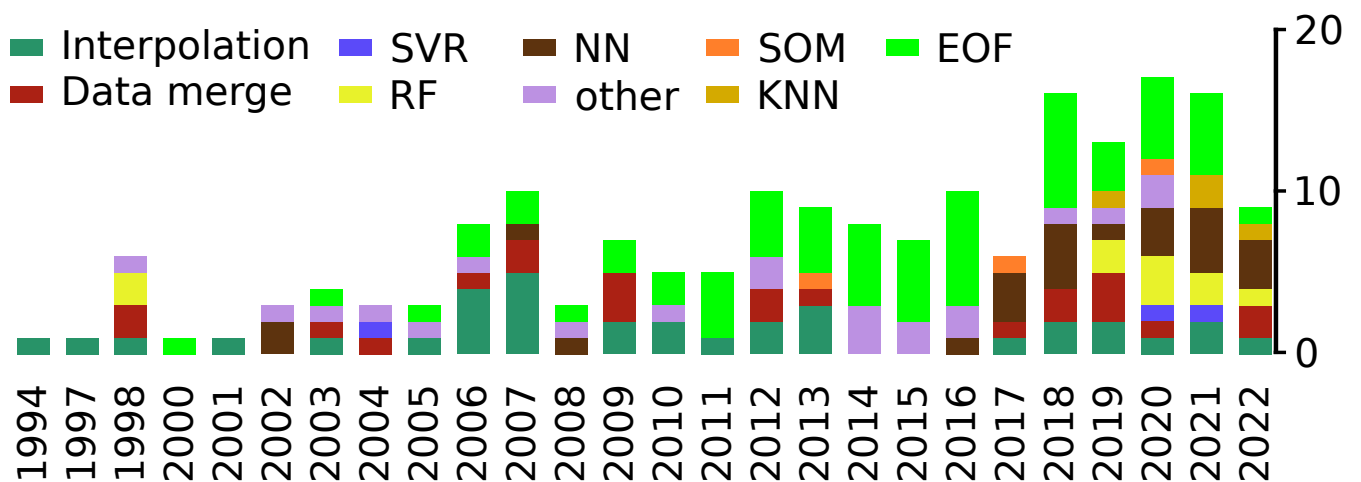
(**a**) Usage of different reconstruction methods throughout the years.
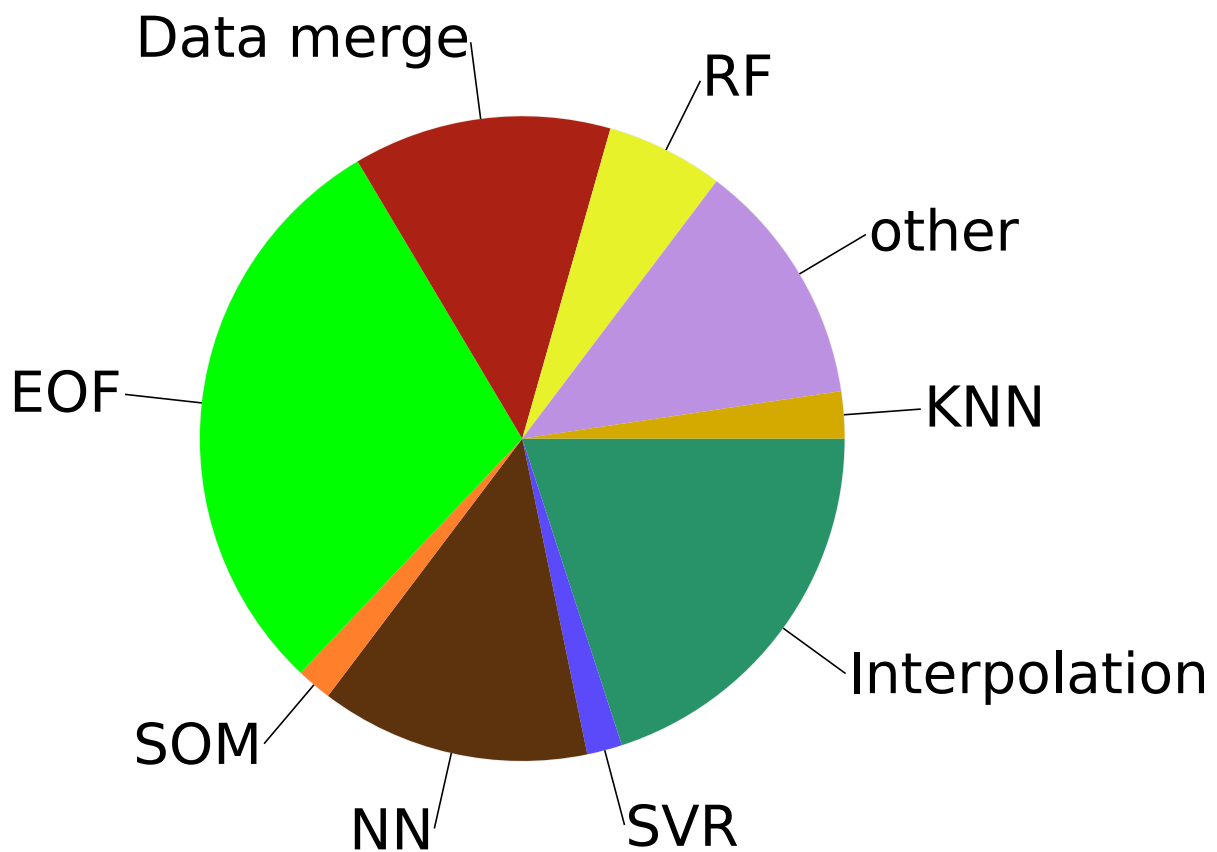
(**b**) General popularity of methods.

Figure 7. Visualisation of the popularity of the methods. While (**a**) showcases the aforementioned trend in popularity, (**b**) displays the overall popularity of methods, with EOF-based methods being the most popular.

Finally, to provide another useful point of view, surveyed articles have been cross-referenced against the three groups of data (chl$_a$, SST and other) and the nine groups of methods (EOF-based, SVR, interpolation-based, KNN, RF, SOM, NN, data merge and other). This cross-referencing is displayed in Table 2.

Table 2. Surveyed articles cross-referenced against the target data and utilised reconstruction method.

	chl$_a$	SST	Other
EOF	[30,34,37,39,43–45,50–52,58,59,63,66,69–74,77–79,81,82,85,87–93,95,133,135,148,157]	[43,44,52,55,87–93,95,105,107,119,120,125,127,128,142–144,150,151,162,191,193–195]	[29–31,33,40,45,50,52,85,87,93,101,102,104,130,148,195]
SVR	[74,84]	[56]	[84]
Interp.	[38,47,57,60,64,67,68,73,76,83,86,132,157]	[47,54,108–113,117,120,121,140,152–154,162,191,193,194]	[54,83,86,99,100,158,163,164]
KNN	[67,74]	n/a	[163,164]
RF	[62,67,74,75,157]	[56,106]	[163]
SOM	[64,157]	n/a	[96]
NN	[59,61,65,70,74,134,137]	[49,56,94,105,106,115,120–124,147,192]	[94,137,139]
Data merge	[35,46,51,68,79,131,148,149]	[46,48,108,113,114,117,152]	[41,42,97,98,148,149,158]
other	[36,53,76,80,92,136,138,155–157]	[92,116,118,126,138,141,145,146,162,192]	[29,32,53,100,103,129,136,138]

5. Discussion and Conclusions

Presented in this paper are more than 130 articles dealing with reconstruction of missing satellite oceanographic data. While this number of articles does not represent the entire effort of the scientific community, it does provide a representative insight into gap-filling methods. The survey provided several results. Firstly, chl$_a$ seems to be the lead targeted type of variable in papers considered, with SST following. Other variables are of significantly lesser interest. From a data-level point of view, level 3 data are the most prominent, with individual swaths of data receiving some attention. Reconstruction at level 1 is significantly hindered as atmospheric correction plays a vital role in satellite observations. Regarding method popularity, DINEOF is undoubtedly the most popular reconstruction method: it is clear that reconstruction efforts in general started garnering ever-increasing attention only after the advent of DINEOF. While DINEOF has proven to be the gold standard, as demonstrated by the abundance in the univariate-time-series box of the novel approach categorisation system, the nearer future might witness a shift from EOF-based to machine-learning-based reconstructions. With the ever-growing and diverging sea of reconstruction methods, the proposed categorisation system might become an invaluable navigation tool amidst the chaos.

Appendix A

For the survey to be valid, the survey itself needs to be bias-free. Since the target of this survey are articles, it is prudent to verify two criteria: both publishing journals and years of publication must be represented as equally as possible. The inclusion of journals is represented in Figure A1, while the inclusion of publication years can be viewed in Figure 7a. Both statistics show that, in general, there is no strong bias towards any single journal or publication year; however, it can be noted that Remote Sensing and Remote Sensing Environment are slightly more popular than the rest of the journals, which should come as no surprise since they specialise in remote sensing and related issues. It is also clear that gap-filling interest rose sometime after 2005, coinciding with the development of DINEOF.
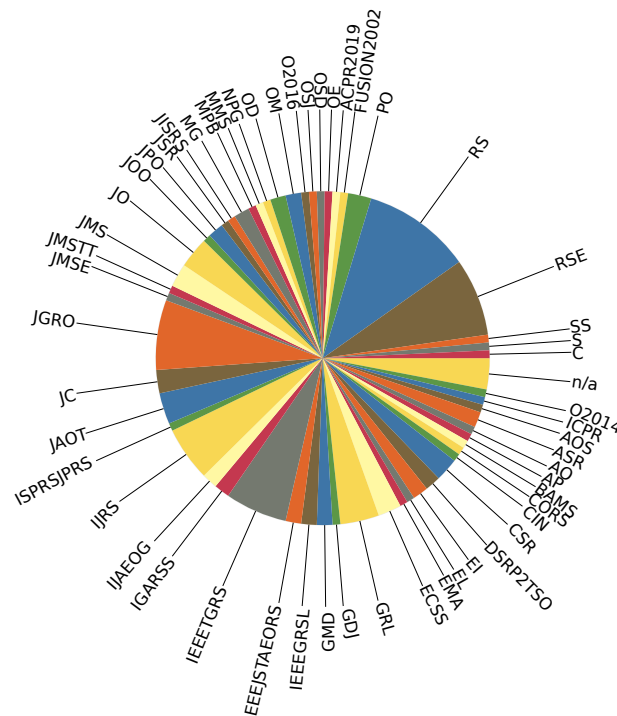
Figure A1. Representation statistics regarding publishing journals of the surveyed articles. List of abbrevations: O2014—2014 Oceans St. John's, OCEANS 2014, ICPR—2018 24th International Conference on Pattern Recognition, AOS—Acta Oceanologica Sinica, ASR—Advances in Space Research, AO—Applied Optics, AP—Aquatic Procedia, BAMS—Bulletin of the American Meteorological Society, CORS—Coastal Ocean Remote Sensing, CIN—Computational Intelligence and Neuroscience, CSR—Continental Shelf Research, DSRP2TSO—Deep Sea Research Part II: Topical Studies in Oceanography, EI—Ecological Indicators, EL—Ecology Letters, EMA—Environmental Monitoring and Assessment, ECSS—Estuarine, Coastal and Shelf Science, GRL—Geophysical Research Letters, GDJ—Geoscience Data Journal, GDM—Geoscientific Model Development, IEEEGRSL—IEEE Geoscience and Remote Sensing Letters, IEEEJSTAEORS—IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEETGRS—Geoscience and Remote Sensing, IEEE Transactions, IGARSS—IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, IJAEOG—International Journal of Applied Earth Observation and Geoinformation, IJRS—International Journal of Remote Sensing, ISPRSJPRS—ISPRS Journal of Photogrammetry and Remote Sensing, JAOT—Journal of Atmospheric and Oceanic Technology, JC—Journal of Climate, JGRO—Journal of Geophysical Research: Oceans, JMSE—Journal of Marine Science and Engineering, JMSTT—Journal of Marine Science and Technology (Taiwan), JMS—Journal of Marine

Systems, JO—Journal of Oceanography, JOO—Journal of Operational Oceanography, JPO—Journal of Physical Oceanography, JSR—Journal of Sea Research, JISRS—Journal of the Indian Society of Remote Sensing, MG—Marine Geodesy, MPB—Marine Pollution Bulletin, MMS—Mediter. Mar. Sci., NPG—Nonlinear Processes in Geophysics, OD—Ocean Dynamics, OM—Ocean Modelling, O2016—OCEANS 2016—Shanghai, OSJ—Ocean Science Journal, OSD—Ocean Science Discussions, OE—Opt. Express, ACPR2019—Pattern Recognition: 5th Asian Conference, ACPR 2019, FUSION2002—Proceedings of the Fifth International Conference on Information Fusion 2002, PO—Progress in Oceanography, RS—Remote Sensing, RSE—Remote Sensing of Environment, S—Science, SS—Spatial Statistics, C—The Cryosphere.

Appendix B

Presented in this Appendix is the Table A1 containing information obtained from over 130 surveyed articles. Information includes: reference, target data type, target data level, spatial and temporal resolution, reconstruction method, proxy data, reported accuracy of the reconstruction and the targeted region (If some information if not available or applicable for a given article, it will be displayed as n/a). Information in Table A1 is taken directly from the cite source. If the article provides multiple accuracies, the best accuracy has been selected. Accuracy measures are displayed as-is, and the definitions can be found in the cited article. Should the reader desire more details, they are referred directly to the source material.

Table A1. Information obtained from the surveyed articles. [1] Names are taken from references where applicable.

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[59]	chl_a	L3	n/a	1 d	VConstruct DINEOF	n/a	**VConstruct**: RMSE = 0.09/0.17 R^2 = 0.44/0.79 **DINEOF**: RMSE = 0.08/0.17 R^2 = 0.54/0.81	Georgia Strait
[72]	chl_a	L3	1 km	8 d	DINEOF	n/a	R^2 = 0.99/0.93	seas around India
[73]	chl_a	L3	4 km	1 d	DINEOF spat.-temp. krig.	n/a	**DINEOF**: RMSE = 0.27 R^2 = 0.99 MAD = 0.02 **spatio-temporal kirging**: RMSE = 0.36 R^2 = 0.88 MAD = 0.15	South China Sea
[74]	chl_a	L3	4 km	8 d	KNN SVR RF NN DINEOF	n/a	**KNN**: RMSE = 0.91 R^2 = 0.79 MAE = 0.45 **SVR**: RMSE = 0.52 R^2 = 0.92 MAE = 0.21 **RF**: RMSE = 0.63 R^2 = 0.89 MAE = 0.22 **NN**: RMSE = 0.68 R^2 = 0.88 MAE = 0.27 **DINEOF**: RMSE = 0.73 R^2 = 0.86 MAE = 0.36	Caspian Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[70]	chl_a	L3	4 km	1 d	DINCAE uni. DICAE multi. DINEOF uni. DINEOF multi.	SST	**DINCAE uni.:** RMSE = 0.26/0.29 R^2 = 0.95/0.91 Bias = −0.03/0.01 Slope = 0.89/0.77 Incpt. = −0.22/−0.57 **DINCEA multi.:** RMSE = 0.25/0.28 R^2 = 0.96/0.91 Bias = −0.02/0.03 Slope = 0.90/0.78 Incpt. = −0.19/−0.53 **DINEOF uni.:** RMSE = 0.28/0.31 R^2 = 0.94/0.89 Bias = −0.01/0.03 Slope = 0.88/0.82 Incpt. = −0.21/−0.43 **DINEOF multi.:** RMSE = 0.32/0.33 R^2 = 0.92/0.88 Bias = −0.01/0.01 Slope = 0.92/0.81 Incpt. = −0.12/−0.47	Luzon Strait seas
[75]	chl_a	L3	9 km	1 d	RF	SST SIC T2M PAR U-V wind bathymetry coord. time	RMSE = 0.16 R^2 = 0.74 MSE = 0.16 MAE = 0.36	Ross Sea
[157]	chl_a	L3	4 km 4 km 9 km 9 km	8 d 8 d 8 d 3 d	lin. temp. int. inv.-dist. weigh. int. ord. krig. spat.-temp. krig. DINEOF SOM ridge reg. (w/, w/o chl_a) RF (w/, w/o chl_a)	SIC SST SLA cloud cover U-V wind	n/a	Beaufort Sea Chukchi Sea Tropical Atlantic Gulf of Mexico
[69]	chl_a	L3	4 km	8 d	DINEOF	n/a	R^2 = 0.96/0.94/0.95	Arabian Sea
[156]	chl_a	L3	n/a	1 d	coverage increase through algorithm refinement	$R_{rs}(443)$ $R_{rs}(555)$ $R_{rs}(670)$	n/a	global
[79]	chl_a	L3	9 km	1 d	DINEOF data merge	n/a	**mean ratio reconstructed/original**: 1.012 ± 0.164 1.015 ± 0.182 0.997 ± 0.287 1.012 ± 0.200	gobal oligothropic ocean global deep waters coastal and inland waters global ocean
[62]	chl_a	L3	1 km	1 d	RF	CI $R_{rs}(469)$ $R_{rs}(555)$ $R_{rs}(645)$ low-qlt chl_a	RMSD = 12.98% R^2 = 0.99 MR = 1.00 UPD = 6.16% MRD = 0.63%	Yellow Sea East China Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[50]	chl_a $R_{rs}(\lambda)$	L2	1 km	1 d	DINEOF VE-DINEOF	n/a	**DINEOF** $\log chl_a$: RMSE = 0.26 S-N = 7.1 **DINEOF** $\log R_{rs}(488)$: RMSE = 0.008 S-N = 9.7 **VE-DINEOF** $\log chl_a$: RMSE = 8×10^{-5} S-N = 1717 **VE-DINEOF** $\log R_{rs}(488)$: RMSE = 4×10^{-5} S-N = 137	Bohai Bay
[81]	chl_a	L3	n/a	1 d	DINEOF	n/a	RE = 0.54%	Arabian Sea
[68]	chl_a	L3	1 km 4 km	1 d	data merge OI	in situ L	**Atlantic merged**: $R^2 = 0.82$ Slope = 0.79 Offset = 0.03 Bias = 0.05 $\sigma = 0.27$ **Atlantic interpolated**: $R^2 = 0.80$ Slope = 0.78 Offset = 0.05 Bias = 0.07 $\sigma = 0.30$ **Global merged**: $R^2 = 0.84$ Slope = 0.87 Offset = 0.01 Bias = 0.01 $\sigma = 0.27$ **Global interpolated**: $R^2 = 0.82$ Slope = 0.82 Offset = 0.02 Bias = 0.02 $\sigma = 0.29$	Atlantic global
[46]	SST chl_a	L2	1 km	1 d	data merge	num. model	**SST**: RMSE = 0.73 bias = -0.003 $R^2 = 0.99$ **chl_a**: RMSE = 1.80 bias = -0.46 $R^2 = 0.75$	Baltic Sea
[67]	chl_a	L3	4 km	1 d	KNN lin. reg. log. reg. DT RF ET	SST SIC PAR T2M U-V wind bathymetry coord. time	poor acc. for the first 4 **RF**: $R^2 = 0.99$ MAE = 0.02 RMSE = 0.06 **ET**: $R^2 = 0.99$ MAE = 0.01 RMSE = 0.04	Ross Sea
[134]	chl_a	L3	4 km	1 d	NN	SST atm. vapor wind rain rate cloud liquid	**full model:** $R^2 = 0.89$ RMSE = 0.30	Korean Peninsula

Table A1. *Cont.*

Ref.	Target Data Type	Data Level	Resolution Spat.	Resolution Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[30]	chl_a K_d $nL(\lambda)$	L1 L3	9 km	1 d	DINEOF	n/a	**mean ratio reconstructed/original**: daily: 1.022 ± 0.29 weekly: 0.983 ± 0.159 monthly: 0.985 ± 0.126	global
[71]	chl_a	L3	4 km	8 d	DINEOF	n/a	RMSE $= 0.41 mg/m^3$	Arabian Sea
[135]	chl_a	L3	1 km	1 d 8 d	DINEOF	n/a	**Year-to-year:** daily: $R^2 = 0.87$ RMSE $= 1.35$ Slope $= 0.84$ Incpt. $= 0.09$ weekly: $R^2 = 0.63$ RMSE $= 1.56$ Slope $= 0.60$ Incpt. $= 0.26$ **Complete span:** daily: $R^2 = 0.91$ RMSE $= 1.29$ Slope $= 0.88$ Incpt. $= 0.09$ weekly: $R^2 = 0.7$ RMSE $= 1.52$ Slope $= 0.67$ Incpt. $= 0.21$	Georgia Strait
[85]	chl_a $R_{rs(\lambda)}$	L3	4.6 km	8 d	DINEOF	n/a	n/a	SW Atlantic
[40]	K_d	L2	0.5 km	1 d	DINEOF	n/a	**mean ratio reconstructed/original**: Bohai Sea: 1.024 ± 0.267 East China Sea: 1.034 ± 0.377	Bohai Sea East China Sea
[87]	chl_a SST PAR	L3	4 km	8 d	DINEOF	n/a	**Error:** SST: < 1 PAR <2 $chl_a < 0.2$	Iceland
[90]	chl_a SST	L3	4 km	1 d	DINEOF multi.	SST chl_a	n/a	Gulf of Mexico
[65]	chl_a	L3	9 km	1 d	NN	SSH SSS SST ARGO *S* (salinity) ARGO *T* time coord.	RMSE $= 0.091$ $R^2 = 0.792$	global
[44]	chl_a SST	L2	n/a	1 d	DINEOF	n/a	n/a	Persian Gulf
[31]	$nL(\lambda)$	L1	4 km (4 × 8 km)	15"	DINEOF	n/a	n/a	North Sea
[63]	chl_a	L3	9 km	1 w	DINEOF	n/a	RMSE $= 0.38$	Red Sea
[43]	chl_a SST	L2	1 km	1 d	DINEOF	n/a	n/a	Peru/Chile
[88]	chl_a SST	L3	4 km	1 d	DINEOF multi.	SST chl_a	RMSE $= 1.57$ $R^2 = 0.97/0.67$ **seasonal cycle on/off**	Georges Bank
[34]	chl_a	L2	1 km	1 d	DINEOF	n/a	n/a	South China Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[36]	chl$_a$	L2	4 km	1 d	scale-changing climatology	low-qlt chl$_a$	n/a	N Atlantic
[82]	chl$_a$	L3	1 km	1 d	DINEOF	n/a	n/a	Sicily Channel
[66]	chl$_a$	L3	n/a	1 m	DINEOF	n/a	RMSE = 0.11 R^2 = 0.88	Bohai Sea Yellow Sea
[78]	chl$_a$	L3	9 km	8 d	O-DINEOF S-DINEOF C-DINEOF D-DINEOF	n/a	**O-DINEOF**: SNR = 3.72 R^2 = 0.97 RMSE = 0.081 MAD = 0.059 **S-DINEOF**: SNR = 4.23 R^2 = 0.97 RMSE = 0.072 MAD = 0.052 **C-DINEOF**: SNR = 3.76 R^2 = 0.97 RMSE = 0.063 MAD = 0.050 **D-DINEOF**: SNR = 5.40 R^2 = 0.98 RMSE = 0.055 MAD = 0.043	Bohai Sea Yellow Sea
[132]	chl$_a$	L3	9 km	8 d	interpolation padding	climatology	n/a	N Atlantic
[91]	chl$_a$ SST	L3	9 km	8 d	DINEOF	n/a	n/a	Gulf of Alaska
[64]	chl$_a$	L3	9 km	1 d	SOM lin. interp.	SST SSH	**daily DINEOF**: RE = 0.25 Kull.dist. = 0.07 **weekly DINEOF**: RE = 0.09 Kull.dist. = 0.008 **daily lin. int.**: RE = 0.33 Kull.dist. = 0.13 **weekly lin. int.**: RE = 0.03 Kull.dist. = 0.002	Mauritania
[39]	chl$_a$	L2	4.5 km	1 d	DINEOF LSEOF RSEOF	n/a	n/a	Galapagos
[92]	chl$_a$ SST	L3	4 km	1 d	DINEOF stat. model	SST SSH	**DINEOF SST**: R^2 = 0.95/0.57 seasonal cycle on/off **DINEOF chl$_a$ / model**: R^2 > 0.92 RMSE < 2.5	Gulf of Mexico
[35]	chl$_a$	L2	1 km 4 km	1 d	data merge	n/a	n/a	California Coast
[77]	chl$_a$	L3	1 km	8 d	DINEOF	n/a	R^2 = 0.82 RMSE = 0.23 BIAS = −0.04	Mediterranean Sea
[58]	chl$_a$	L3	9 km	1 d	DINEOF	n/a	n/a	Peru–Chile
[138]	chl$_a$ SST MLD	L3	4.5 km	8 d	GSM01 semi-anal. OC (Ocean Colour) model	n/a	n/a	Gulf of Cadiz

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[86]	chl$_a$ PAR	L3	9 km	1 w	spat.-temp. int.	n/a	n/a	global
[52]	chl$_a$ TSM SST	L2	1 km	1 d	DINEOF	n/a	**TMS min-max**: $R^2 = 0.45 - 0.95$ $S - N = 1 - 2.7$ **chl$_a$ min-max**: $R^2 = 0.12 - 0.70$ $S - N = 0.88 - 1.27$	English Channel
[195]	TSM SST TSM wind	n/a	n/a	n/a	DINEOF uni. DINEOF multi.	wind chl$_a$ SST	n/a	English Channel Gulf of Mexico Mediterranean Sea Black Sea
[57]	chl$_a$	L2	1 km	1 d	kriging	n/a	$R^2 = 0.72$ (log)	Ireland Biscay Bay Iberia
[95]	chl$_a$ SST	L3	4 km	1 d	DINEOF mulit.	chl$_a$ SST	**DINEOF SST**: $R^2 = 0.93/0.53$ **seasonal cycle on/off**	S Atlantic Bight
[194]	SST	n/a	1 km	1 d	DINEOF DINEOF Pruning OI	n/a	**DINEOF**: $MSE = 0.086$ **DINEOF pruning**: $MSE = 0.055$ **OI**: $MSE = 0.627$	Tanganyika Lake
[53]	chl$_a$ CDM b_{bp}	L2	1 km	1 d	GSM01 semi-anal. OC model	$nL(412)$ $nL(443)$ $nL(490)$ $nL(510)$ $nL(531)$ $nL(547)$ $nL(557)$ $nL(620)$ $nL(671)$	**chl$_a$** : $R^2 = 0.678$ $RMSE = 0.399$ $BIAS = -0.145$ **CDM**: $R^2 = 0.465$ $RMSE = 0.372$ $BIAS = 0.0128$ **b$_{bp}$** : $R^2 = 0.503$ $RMSE = 0.197$ $BIAS = -0.053$	global
[42]	$nL(\lambda)$	L2	2 km	1 d	data merge	n/a	*nL(412)*: $R^2 = 0.48$ $RMSD = 0.233$ *nL(440)*: $R^2 = 0.72$ $RMSD = 0.210$ *nL(490/500)*: $R^2 = 0.86$ $RMSD = 0.213$ *nL(551/555)*: $R^2 = 0.84$ $RMSD = 0.207$ *nL(667/674)*: $R^2 = 0.73$ $RMSD = 0.05$	Mediterranean Sea
[80]	chl$_a$	L3	1/24°	1 d	turbulent cascading	n/a	n/a	global
[89]	chl$_a$ SST	L3	4 km	1 d	DINEOF multi.	SST chl$_a$	**DINEOF SST**: $R^2 = 0.97/0.69$ seasonal cycle on/off	S Atlantic Bight
[60]	chl$_a$	L3	n/a	1 m	spat. int.	n/a	n/a	N Atlantic
[37]	chl$_a$	L2	n/a	1 d	DINEOF	n/a	$R^2 = 0.912$ $RMSE = 0.648$	Adriatic Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[83]	chl_a a_{ph} b_{bp}	L3	1 km	1 d	spat.-temp. int.	n/a	n/a	Mississippi Bight
[41]	$nL(\lambda)$	L2	1 km	1 d	data merge	n/a	**nL(412)**: R^2 = 0.77 RMSD = 0.203 **nL(440)**: R^2 = 0.87 RMSD = 0.179 **nL(500)**: R^2 = 0.93 RMSD = 0.157 **nL(555)**: R^2 = 0.93 RMSD = 0.132 **nL(675)**: R^2 = 0.89 RMSD = 0.026	Adriatic Sea
[38]	chl_a	L2	1.5 km	1 d	ord. krig.	n/a	n/a	North Sea
[93]	chl_a SST wind	L3	1 km	1 d	DINEOF uni. DINEOF multi.	chl_a SST wind	**chl$_a$**: **SST**: RMSE = 0.76 R^2 = 0.59 **SST + wind**: RMSE = 0.75 R^2 = 0.60 **SST + chl$_a$**: RMSE = 0.62 R^2 = 0.71 **SST + day lag SST + chl$_a$**: RMSE = 0.6 R^2 = 0.69 **wind RMSE** = 2.8	West Florida Shelf
[76]	chl_a	L3	1 km	1 d	weigh. avg. OI	n/a	**Weighted averaging**: R^2 = 0.47 RMS = 4.62 BIAS = 0.68 **Optimal interpolation**: R^2 = 0.47 RMS = 4.62 BIAS = 0.68	Atlantic Ocean
[136]	chl_a a_{ph} a_{cdm} b_{bp}	L3	4.5 km 1 km	1 d	GSM01 semi-anal. OC model	$nL(443)$	R^2 = 0.943 RMSE = 0.178	global
[131]	chl_a	L3	n/a	1 m	data merge	in situ	n/a	global
[61]	chl_a	L3	1 km	1 d	CCGAN	SST SSH low- res. chl_a	SSIM = 0.9462 MSE = 0.004 RE = 0.039	Adriatic Sea
[51]	chl_a	L2	1 km	1 d	data merge DINEOF	n/a	**Data merge**: RMSE = 0.408 **DINEOF**: RMSE = 0.38	Mexico
[109]	SST	L3	1 °	1 d	OI	n/a	Data-to-guess-error: 1.25	global
[193]	SST	n/a	1.3 km	1 d	DINEOF OI	n/a	**DINEOF**: cloud coverage: 40% RMSE = 0.89 60% **OI**: similar RMSE = 0.78 80% RMSE = 1.25	Adriatic Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[107]	SST	L3	18 km	8 d	DINEOF	n/a	n/a	Japan East Sea
[191]	SST	n/a	4 km	1 d	DINEOF OI	n/a	**DINEOF:** RMSE = 0.42	Corsica
[111]	SST	L3	1.2 km 2 km	1 d	OI	n/a	RMSE = 0.69–0.82	Baltic Sea North Sea
[140]	SST	L3	1.2 km 1.5 km	1 d	OI	n/a	RMSE = 0.78	Baltic Sea North Sea
[123]	SST	L3	4 km	1 m	NN	MSLP T2M wind CC dewpoint T	RMSE = 0.1–0.7	Mediterranean Sea
[108]	SST	L3	4.5 km	1 d	OI data merge	SIC	N-S ratio: 0.5-0.35	global
[119]	SST	L3	1 km	1 d	DINEOF	n/a	R^2 = 0.98 mean diff: 0.243	Gulf of Trieste
[128]	SST	L3	4.8 km	1 d	DINEOF	n/a	RMSE = 0.46	Black Sea
[47]	SST chl_a	L2	1 km	n/a	OI	n/a	**SST:** RE = 1.2% ABS = 0.29 **chl_a:** RE = 23% ABS = 0.56	Black Sea
[127]	SST	L3	1°	5 d	DINEOF EOF	n/a	**DINEOF:** RMSE = 0.91 **EOF:** RMSE = 0.82	Yangtze River estuary
[110]	SST	L3	5 km	1 d	kriging	n/a	BIAS = -0.285 STD = 0.68	English Channel
[130]	TSM	L3	n/a	1 d	DINEOF	n/a	RE = 40% RMSE = 7.36 R^2 = 0.75	English Channel
[55]	SST	L2	1 km	1 d	DINEOF	n/a	RMSE = 0.265 MAD = 0.396 R^2 = 0.635	Bay of Biscay
[158]	SSH	L3	2 "	1 d	OI data merge	n/a	RMSE = 2	global
[104]	SSH	L3	0.25°	1 w	EOF	in situ	**Pacific:** mean R^2 = 0.42 median R^2 = 0.44 **Global:** mean R^2 = 0.48 median R^2 = 0.49	Pacific Ocean global
[54]	SST SSS	L2	1 km	1 d	kriging	n/a	**SST:** ME = 0.3 MAE = 0.65 RMSE = 0.84 R^2 = 0.99 **SSS:** ME = 0.85 MAE = 2.08 RMSE = 2.76 R^2 = 0.82	Chesapeake Bay
[144]	SST	L3	n/a	1 d	DINEOF	n/a	MAD $\leq$ 1	Mediterranean Sea
[101]	SSH	L3	n/a	n/a	EOF	in situ	n/a	global

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[125]	SST	L3	4 km	1 m	DINEOF I-DINEOF	n/a	**DINEOF**: $R^2 = 0.9925$ S-N = 9.7743 RMSE = 0.3144 MAD = 0.1680 **I-DINEOF**: $R^2 = 0.9983$ S-N = 25.4548 RMSE = 0.1526 MAD = 0.1140	South China Sea
[142]	SST	L3	4 km	1 d	DINEOF I-DINEOF VE-DINEOF	n/a	**DINEOF**: $R^2 = 0.9943$ S-N = 11.0682 RMSE = 0.2773 MAD = 0.1515 **I-DINEOF**: $R^2 = 0.9964$ S-N = 17.1149 RMSE = 0.2215 MAD = 0.1631 **VE-DINEOF**: $R^2 = 0.9987$ S-N = 19.9641 RMSE = 0.1303 MAD = 0.0155	South China Sea
[143]	SST	L3	4 km	1 d	DINEOF	n/a	BIAS = −0.34 RMSE = 0.37 $R^2 = 0.95$	South China Sea
[33]	SSS	L2	0.15°	1 d	DINEOF	n/a	RMSE = 0.66 CMSE = 0.63 BIAS = 0.19 $R^2 = 0.73$	N-E Atlantic Ocean Mediterranean Sea
[147]	SST	L3	0.25°	1 d	NN	n/a	**1 day**: $R^2 = 0.85$ RMSE = 0.62 MAE = 0.53 **2 days**: $R^2 = 0.84$ RMSE = 0.63 MAE = 0.54 **3 days**: $R^2 = 0.84$ RMSE = 0.65 MAE = 0.55 **4 days**: $R^2 = 0.82$ RMSE = 0.67 MAE = 0.56 **5 days**: $R^2 = 0.81$ RMSE = 0.69 MAE = 0.58	Indian Ocean
[96]	sub. vel.	L3	0.25°	1 d	SOM	SSH SST ARGO vel.	$R^2 = 0.956$ RMSE = 2.8 MAE = 18°	Antarctic Ocean
[133]	chl$_a$	L3	4 km	1 d	DINEOF	n/a	**MOIDS**: RMSE = 0.38 $R^2 = 0.74$ **VIIRS**: RMSE = 0.13 $R^2 = 0.83$	Laizhou Bay

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[192]	SST	n/a	n/a	n/a	GAN Telea TV Patch-Match	n/a	**GAN-1**: PNAR = 0.1310 **GAN-2**: PNAR = 0.0862 **Telea**: PNAR = 0.0333 **TV**: PNAR = 0.166 **Patch-Match**: PNAR = 0.1085	n/a
[122]	SST	L3	2 km	1 d	GAN	n/a	**avg. occlusion MSE**: $Loss_1 = 0.222$ $Loss_2 = 0.121$ $Loss_3 = 0.166$ $Loss_4 = 0.121$ $Loss_5 = 0.107$ $Loss_6 = 0.193$ $Loss_7 = 0.112$	Pacific Ocean
[115]	SST	L3	2 km	1 d	GAN	n/a	**avg. occlusion MSE**: $Loss_1 = 0.120$ $Loss_2 = 0.136$ $Loss_3 = 0.132$ $Loss_4 = 0.169$	N Pacific
[102]	SSH	L3	0.25°	1 w	CSEOF	SST SLP	$R^2 = 0.79$	Indian Ocean
[105]	SST	L3	4 km	1 d	DINCAE DINEOF	n/a	**DINCAE**: RMSE = 1.1362 CRMSE = 1.0879 BIAS = -0.3278 **DINEOF**: RMSE = 1.1676 CRMSE = 1.1102 BIAS = −0.3616	Provencal basin
[155]	chl_a	L3	4 km	8 d	gridfill	n/a	$R^2 = 0.968$ RMSE = 0.0586	Yellow Sea
[56]	SST	L2	1 km	1 d	NN SVR RF	in situ coord. time cloud cover	**NN**: $R^2 = 0.77$ RMSE = 0.91 MAE = 0.69 **SVR**: $R^2 = 0.80$ RMSE = 0.79 MAE = 0.59 **RF**: $R^2 = 0.78$ RMSE = 0.85 MAE = 0.64	Arabian Sea Bay of Bengal
[124]	SST	L3	1 km	1 d	GAN	n/a	AVG = 0.35 SVD = 0.67 MAE = 0.80 RMSPE = 10.36	Yellow Sea
[163]	wind	L4	0.125°	6 h	lin. reg. KNN DT NN	n/a	**lin. reg.**: Amp = 0.30 ± 0.39 Ang = 0.40 ± 8.74 **KNN**: Amp = 0.78 ± 0.67 Ang = 10.10 ± 17.54 **DT**: Amp = 0.85 ± 0.82 Ang = 11.12 ± 18.79 **NN**: Amp = 0.41 ± 0.46 Ang = 5.22 ± 10.58	Adriatic Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[164]	wind	L4	0.5°	6 h	lin. reg. KNN	n/a	**lin. reg.:** Amp =2.16 ± 1.55 Ang = 24.49 ± 30.75 **KNN:** Amp = 2.19 ± 1.59 Ang = 24.70 ± 31.98	Mediterranean Sea
[148]	chl_a K_d SPM	L3	750 m 300 m	1 d	data merge DINEOF	n/a	**chl_a:** Mean ratio = 1.016 ± 0.222 Mean diff = -0.013 ± 0.543 **K_d:** Mean ratio = 1.012 ± 0.154 Mean diff = −0.008 ± 0.165 **SPM:** Mean ratio = 1.020 ± 0.205 Mean diff = −0.17 ± 0.485	global
[106]	SST	L3	4 km 10 km	1 d	DINCAE RF merge	coord. time	**DINCAE:** $R^2 = 0.99$ BIAS = 0.02 RMSE = 0.75 rRMSE = 4.23 MAE = 0.55 **RF:** $R^2 = 0.99$ BIAS = 0.02 RMSE = 0.87 rRMSE = 4.41 MAE = 0.61	NW Pacific
[94]	SST SSH	L3	4 km n/a (non-gridded)	1 d	DINCAE	SST wind chl_a	**SST:** RMSE = 0.54 10th–90th = 0.05–0.82 **SSH:** n/a	Adriatic Sea Mediterranean Sea
[126]	SST	L3	1/20°	1 d	Hidden Markov Model	n/a	n/a	China coastal waters
[129]	TIT (thin-ice thickness)	L3	1 km	1 d	weig. feat. recon.	n/a	$R^2 = 0.81$ RMSE = 2140	Brunt Ice Shelf
[48]	SST	L2	n/a	1 m	data merge	n/a	$R^2 = 0.97$ RMSE = 0.91 BIAS = -0.14	Gulf of Mexico
[154]	SST	L3	1 km 1/4°	1 d	OI	n/a	**seas. cycle:** $R^2 = 0.98/0.81$ RMSE = 0.69/0.61	Florida Bay
[151]	SST	L3	n/a	1 d	DINEOF	n/a	n/a	Adriatic Sea
[84]	chl_a $nL(\lambda)$	L3	4.5 km 9 km	1 d	SVR merge	n/a	n/a	global
[137]	chl_a $nL(\lambda)$ $K_d(\lambda)$	L3	4.5 km 9 km	1 d	NN merge	n/a	n/a	global
[97]	OC	L3	n/a	1 d	data merge	n/a	58% coverage increase	global
[98]	OC	L3	n/a	1 d	data merge	n/a	44% coverage increase	global
[153]	SST	L3	2.3 km	1 d	interpolation	n/a	n/a	NW Atlantic
[100]	SLA	L3	6 km	10 d	interpolation multiscale est.	n/a	RMSE = 4.90 ± 3.11	Mediterranean Sea

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[45]	chl$_a$ TSM	L2	1 km	1 d	DINEOF	n/a	**chl$_a$**: $R^2 = 0.94$ RMSE = 0.13 **TSM**: $R^2 = 0.93$ RMSE = 0.05	Ariake Sea
[162]	SST	L4	0.05 ° 0.25 °	1 d	OI DINEOF analog data ass.	n/a	**OI**: RMSE = 0.45 ± 0.08 $R^2 = 0.76 ± 0.07$ **DINEOF**: RMSE = 0.40 ± 0.07 $R^2 = 0.83 ± 0.05$ **data ass.**: RMSE = 0.31 ± 0.05 $R^2 = 0.88 ± 0.02$	South Africa
[120]	SST	L3	1/20°	1 d	OI analog data ass. DINEOF NN	n/a	**OI**: RMSE = 0.75 **analog data ass.**: RMSE = 0.45 **DINEOF**: RMSE = 0.54 **NN**: RMSE = 0.43	South Africa
[139]	SSH	L3	0.2°	10 d	NN	n/a	RMSE = 10–45 $R^2 = 0.99$–0.45	Mediterranean Sea
[112]	SST	L3	1.4 km	1 d	OI	n/a	n/a	Gulf of Maine
[103]	SSH	L3	2.5°	1 d	weigh. filter	n/a	n/a	global
[113]	SST	L3	6 km 2 km 1.6 km 25 km	1 d	OI data merge	in situ	BIAS = 0.17 RMSE = 0.81	California
[150]	SST	L3	1.1 km	1 m	EOF forecast	n/a	n/a	Alboran Sea
[121]	SST	L3	9 km	1 d	NN interpolation	n/a	**NN**: BIAS = −0.28 RMSE = 0.87 $R^2 = 0.86$ **interpolation**: BIAS = −0.49 RMSE = 1.24 $R^2 = 0.78$	W Mediterranean
[117]	SST	L3	0.088° 0.125° 0.167° 0.25°	1 d	OI data merge	n/a	n/a	global
[146]	SST	L3	0.5°	1 d	corr. formula	SIC climat. *S*	n/a	N Atlantic
[116]	SST	L3	0.2°	1 d	Kalman filter	n/a	n/a	N Atlantic
[114]	SST	L3	0.01° 0.25°	1 d	data merge	n/a	BIAS = −0.01 STD = 0.95	Pacific Ocean
[32]	LSI (landfast sea-ice)	L1	1 km	1 d	temp. comp.	n/a	n/a	Mertz Glacier Tongue
[152]	SST	L3	1 km 5 km 9 km 25 km	1 d	data merge interpolation	SIC	RMSE = 0.3 BIAS = −0.2	global

Table A1. *Cont.*

Ref.	Target Data Type	Level	Resolution Spat.	Temp.	Method	Proxy Data	Reported Accuracy	Area of Interest [1]
[149]	chl_a K_d	L3	n/a	1 d	data merge	R_{rs} nL	chl_a ratio: 1 ± 0.22 K_d ratio: 1.05 ± 0.15	global
[29]	SSS	L1 L3	0.05° 0.25°	1 d 9 d	Non-Bayesian retrieval DINEOF multifractal fusion	SST B_λ	**N.B.R.:** RMSE = 0.35 **DINEOF:** RMSE = 0.54 **M.F.:** RMSE = 0.32	N Atlantic Mediterranean Sea
[49]	SST	L2	1 km	1 d	NN	nR	RMSE = 4.46 $R^2 = 0.94$	seas around USA
[141]	SST	L3	0.25°	1 d	Hidden Markov Model	n/a	n/a	South Africa
[118]	SST	L3	0.05° 0.25°	1 d	Kelman filter	low. res. SST	n/a	Malvinas
[99]	SIC	L3	1 km	1 d	spat. scaling	n/a	n/a	Svalbard archipelago
[145]	SST	L3	0.05° 0.25°	1 d	semivariogram model	low. res. SST	n/a	Malvinas

References

1. Belward, A.; Bourassa, M.; Dowell, M.; briggs, S.; Dolman, H.A.; Holmlund, K.; Husband, R.; Quegan, S.; Simmons, A.; Sloyan, B.; et al. The Global Observing System for Climate: Implementation Needs. 2016. Available online: https://public.wmo.int/en/resources/library/global-observing-system-climate-implementation-needs (accessed on 27 December 2022)
2. IOCCG. *Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology*; IOCCG: Dartmouth, NS, Canada, 2008. [CrossRef]
3. Bojinski, S.; Verstraete, M.; Peterson, T.C.; Richter, C.; Simmons, A.; Zemp, M. The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 1431–1443. [CrossRef]
4. Steele, J.; Thorpe, S.; Turekian, K. *Measurement Techniques, Platforms & Sensors: A Derivative of the Encyclopedia of Ocean Sciences*; Elsevier: Chicago, IL, USA, 2009.
5. Campbell, J.B.; Wynne, R.H. *Introduction to Remote Sensing*; Guilford Press: New York, NY, USA, 2011.
6. Robinson, I.S. *Discovering the Ocean from Space: The Unique Applications of Satellite Oceanography*; Springer: Berlin, Germany, 2016.
7. Gordon, H. *Physical Principles of Ocean Color Remote Sensing*; University of Miami: Coral Gables, FL, USA, 2019. [CrossRef]
8. Coble, P. Marine Optical Biogeochemistry: The Chemistry of Ocean Color. *Chem. Rev.* **2007**, *107*, 402–418. [CrossRef] [PubMed]
9. Strong, A.E.; McClain, E.P. Improved Ocean Surface Temperatures From Space—Comparisons With Drifting Buoys. *Bull. Am. Meteorol. Soc.* **1984**, *65*, 138–142. [CrossRef]
10. McClain, E.P. Global sea surface temperatures and cloud clearing for aerosol optical depth estimates. *Int. J. Remote Sens.* **1989**, *10*, 763–769.
11. Joseph, G. *Fundamentals of Remote Sensing*, 2nd ed.; Universities Press: Hyderabad, India, 2005.
12. Wylie, D.; Jackson, D.L.; Menzel, W.P.; Bates, J.J. Trends in Global Cloud Cover in Two Decades of HIRS Observations. *J. Clim.* **2005**, *18*, 3021–3031. [CrossRef]
13. Wentz, F.J.; Gentemann, C.; Smith, D.; Chelton, D. Satellite Measurements of Sea Surface Temperature Through Clouds. *Science* **2000**, *288*, 847–850. [CrossRef]
14. Ackerman, S.A.; Strabala, K.I.; Menzel, W.P.; Frey, R.A.; Moeller, C.C.; Gumley, L.E. Discriminating clear sky from clouds with Modis. *J. Geophys. Res. Atmos.* **1998**, *103*, 32141–32157. [CrossRef]
15. Wang, M.; Shi, W. Cloud Masking for Ocean Color Data Processing in the Coastal Regions. *Geosci. Remote Sens. IEEE Trans.* **2006**, *44*, 3105–3196. [CrossRef]
16. Wang, M.; Bailey, S.W. Correction of sun glint contamination on the SeaWiFS ocean and atmosphere products. *Appl. Opt.* **2001**, *40*, 4790–4798. [CrossRef]
17. Reynolds, R.W.; Folland, C.K.; Parker, D.E. Biases in satellite-derived sea-surface-temperature data. *Nature* **1989**, *341*, 728–731. [CrossRef]
18. Reynolds, R.W. Impact of Mount Pinatubo Aerosols on Satellite-derived Sea Surface Temperatures. *J. Clim.* **1993**, *6*, 768–774. [CrossRef]

19. Sathyendranath, S.; Brewin, R.J.; Brockmann, C.; Brotas, V.; Calton, B.; Chuprin, A.; Cipollini, P.; Couto, A.B.; Dingle, J.; Doerffer, R.; et al. An Ocean-Colour Time Series for Use in Climate Studies: The Experience of the Ocean-Colour Climate Change Initiative (OC-CCI). *Sensors* **2019**, *19*, 4285. [CrossRef]

20. Meister, G.; McClain, C.R. Point-spread function of the ocean color bands of the Moderate Resolution Imaging Spectroradiometer on Aqua. *Appl. Opt.* **2010**, *49*, 6276–6285. [CrossRef]

21. Varnai, T.; Marshak, A. Effect of Cloud Fraction on Near-Cloud Aerosol Behavior in the MODIS Atmospheric Correction Ocean Color Product. *Remote Sens.* **2015**, *7*, 5283–5299. [CrossRef]

22. Hu, C.; Feng, L.; Lee, Z.; Davis, C.O.; Mannino, A.; McClain, C.R.; Franz, B.A. Dynamic range and sensitivity requirements of satellite ocean color sensors: Learning from the past. *Appl. Opt.* **2012**, *51*, 6045–6062. [CrossRef]

23. NASA Ocean Color–SST. Available online: https://oceancolor.gsfc.nasa.gov/docs/modis_sst/ (accessed on 27 December 2022).

24. Rajeesh, R.; Dwarakish, G. Satellite Oceanography—A review. *Aquat. Procedia* **2015**, *4*, 165–172. [CrossRef]

25. Traon, P.Y.L.; Antoine, D.; Bentamy, A.; Bonekamp, H.; Breivik, L.; Chapron, B.; Corlett, G.; Dibarboure, G.; DiGiacomo, P.; Donlon, C.; et al. Use of satellite observations for operational oceanography: Recent achievements and future prospects. *J. Oper. Oceanogr.* **2015**, *8*, s12–s27.

26. NASA Ocean Color. Available online: https://oceancolor.gsfc.nasa.gov/ (accessed on 27 December 2022).

27. Monitoring the Weather and Climate from Space. Available online: https://www.eumetsat.int/ (accessed on 27 December 2022).

28. What's behind the curtain of the NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission? Training Activity. Available online: https://www.us-ocb.org/wp-content/uploads/sites/43/2022/08/L14_PACE_Applications_PACECLASS_20220805.pdf/ (accessed on 27 December 2022).

29. Olmedo, E.; Taupier-Letage, I.; Turiel, A.; Alvera-Azcárate, A. Improving SMOS Sea Surface Salinity in the Western Mediterranean Sea through Multivariate and Multifractal Analysis. *Remote Sens.* **2018**, *10*, 485. [CrossRef]

30. Liu, X.; Wang, M. Gap Filling of Missing Data for VIIRS Global Ocean Color Products Using the DINEOF Method. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4464–4476. [CrossRef]

31. Alvera-Azcárate, A.; Vanhellemont, Q.; Ruddick, K.; Barth, A.; Beckers, J.M. Analysis of high frequency geostationary ocean colour data using DINEOF. *Estuar. Coast. Shelf Sci.* **2015**, *159*, 28–36. [CrossRef]

32. Fraser, A.; Massom, R.; Michael, K. A Method for Compositing Polar MODIS Satellite Images to Remove Cloud Cover for Landfast Sea-Ice Detection. *Geosci. Remote Sens. IEEE Trans.* **2009**, *47*, 3272–3282. [CrossRef]

33. Alvera-Azcárate, A.; Barth, A.; Parard, G.; Beckers, J.M. Analysis of SMOS sea surface salinity data using DINEOF. *Remote Sens. Environ.* **2016**, *180*, 137–145. [CrossRef]

34. Liu, M.; Liu, X.; Ma, A.; Li, T.; Du, Z. Spatio-temporal stability and abnormality of chlorophyll-a in the Northern South China Sea during 2002–2012 from MODIS images using wavelet analysis. *Cont. Shelf Res.* **2014**, *75*, 15–27. [CrossRef]

35. Kahru, M.; Kudela, R.M.; Manzano-Sarabia, M.; Greg Mitchell, B. Trends in the surface chlorophyll of the California Current: Merging data from multiple ocean color satellites. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2012**, *77–80*, 89–98. [CrossRef]

36. Land, P.; Shutler, J.; Platt, T.; Racault, M. A novel method to retrieve oceanic phytoplankton phenology from satellite data in the presence of data gaps. *Ecol. Indic.* **2014**, *37*, 67–80. [CrossRef]

37. Mauri, E.; Poulain, P.M.; Južnič-Zonta, V. MODIS chlorophyll variability in the northern Adriatic Sea and relationship with forcing parameters. *J. Geophys. Res. Ocean.* **2007**, *112*. [CrossRef]

38. Müller, D. *Estimation of Algae Concentration in Cloud Covered Scenes Using Geostatistical Methods*; GKSS: Geesthacht, Germany, 2007.

39. Taylor, M.H.; Losch, M.; Wenzel, M.; Schröter, J. On the Sensitivity of Field Reconstruction and Prediction Using Empirical Orthogonal Functions Derived from Gappy Data. *J. Clim.* **2013**, *26*, 9194–9205. [CrossRef]

40. Liu, X.; Wang, M. Analysis of ocean diurnal variations from the Korean Geostationary Ocean Color Imager measurements using the DINEOF method. *Estuar. Coast. Shelf Sci.* **2016**, *180*, 230–241. [CrossRef]

41. Mélin, F.; Zibordi, G. Optically based technique for producing merged spectra of water-leaving radiances from ocean color remote sensing. *Appl. Opt.* **2007**, *46*, 3856–3869. [CrossRef]

42. Mélin, F.; Zibordi, G.; Djavidnia, S. Merged series of normalized water leaving radiances obtained from multiple satellite missions for the Mediterranean Sea. *Adv. Space Res.* **2009**, *43*, 423–437. [CrossRef]

43. Corredor Acosta, A.; Morales, C.; Hormazabal, S.; Andrade, I.; Correa-Ramirez, M. Phytoplankton phenology in the coastal upwelling region off central-southern Chile (35 °S–38 °S): Time-space variability, coupling to environmental factors, and sources of uncertainty in the estimates. *J. Geophys. Res. Ocean.* **2015**, *120*, 813–831. [CrossRef]

44. Moradi, M.; Kabiri, K. Spatio-temporal variability of SST and Chlorophyll-a from MODIS data in the Persian Gulf. *Mar. Pollut. Bull.* **2015**, *98*, 14–25. [CrossRef]

45. Yang, M.; Khan, F.A.; Tian, H.; Liu, Q. Analysis of the Monthly and Spring-Neap Tidal Variability of Satellite Chlorophyll-a and Total Suspended Matter in a Turbid Coastal Ocean Using the DINEOF Method. *Remote Sens.* **2021**, *13*, 632. [CrossRef]

46. Konik, M.; Kowalewski, M.; Bradtke, K.; Darecki, M. The operational method of filling information gaps in satellite imagery using numerical models. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *75*, 68–82. [CrossRef]

47. Pukhtyar, L.D.; Stanichny, S.V.; Timchenko, I.E. Optimal interpolation of the data of remote sensing of the sea surface. *Phys. Oceanogr.* **2009**, *19*, 225. [CrossRef]

48. Barnes, B.B.; Hu, C. A Hybrid Cloud Detection Algorithm to Improve MODIS Sea Surface Temperature Data Quality and Coverage Over the Eastern Gulf of Mexico. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3273–3285. [CrossRef]

49. Wang, J.; Deng, Z. Development of MODIS data-based algorithm for retrieving sea surface temperature in coastal waters. *Environ. Monit. Assess.* **2017**, *189*, 286. [CrossRef]

50. Ping, B.; Meng, Y. Reconstruct Oceanic Chlorophyll and Reflectance Data Based on an Improved VE-DINEOF Method. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7984–7987. [CrossRef]

51. Lomelı-Huerta, J.R.; Rivera-Caicedo, J.P.; De-la Torre, M.; Acevedo-Juárez, B.; Cepeda-Morales, J.; Avila-George, H. An approach to fill in missing data from satellite imagery using data-intensive computing and DINEOF. *PeerJ Comput. Sci.* **2022**, *8*, e979. [CrossRef]

52. Sirjacobs, D.; Alvera-Azcárate, A.; Barth, A.; Lacroix, G.; Park, Y.; Nechad, B.; Ruddick, K.; Beckers, J.M. Cloud filling of ocean colour and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology. *J. Sea Res.* **2011**, *65*, 114–130. [CrossRef]

53. Maritorena, S.; d'Andon, O.H.F.; Mangin, A.; Siegel, D.A. Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sens. Environ.* **2010**, *114*, 1791–1804. [CrossRef]

54. Urquhart, E.A.; Hoffman, M.J.; Murphy, R.R.; Zaitchik, B.F. Geospatial interpolation of MODIS-derived salinity and temperature in the Chesapeake Bay. *Remote Sens. Environ.* **2013**, *135*, 167–177. [CrossRef]

55. Ganzedo, U.; Alvera-Azcárate, A.; Esnaola, G.; Ezcurra, A.; Sáenz, J. Reconstruction of sea surface temperature by means of DINEOF: A case study during the fishing season in the Bay of Biscay. *Int. J. Remote Sens.* **2011**, *32*, 933–950. [CrossRef]

56. Sunder, S.; Ramsankaran, R.; Ramakrishnan, B. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 228–240. [CrossRef]

57. Saulquin, B.; Gohin, F.; Garrello, R. Regional Objective Analysis for Merging High-Resolution MERIS, MODIS/Aqua, and SeaWiFS Chlorophyll-a Data From 1998 to 2008 on the European Atlantic Shelf. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 143–154. [CrossRef]

58. Correa-Ramirez, M.A.; Hormazabal, S.E.; Morales, C.E. Spatial patterns of annual and interannual surface chlorophyll-a variability in the Peru–Chile Current System. *Prog. Oceanogr.* **2012**, *92–95*, 8–17. [CrossRef]

59. Ehrler, M.; Ernst, N. VConstruct: Filling Gaps in Chl-a Data Using a Variational Autoencoder, *arXiv* **2021**, arXiv:2101.10260. https://doi.org/10.48550/ARXIV.2101.10260.

60. Beaugrand, G.; Edwards, M.; Brander, K.; Luczak, C.; Ibanez, F. Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic. *Ecol. Lett.* **2008**, *11*, 1157–1168. [CrossRef]

61. Ćatipović, L.; Kalinić, H.; Županović, T.; Sathyendranath, S.; Dingle, J.; Jackson, T.; Matić, F. Implementation of GAN-based satellite derived chlorophyll-a concentration gap reconstruction. 2023, *manuscript submitted for publication.*

62. Chen, S.; Hu, C.; Barnes, B.B.; Xie, Y.; Lin, G.; Qiu, Z. Improving ocean color data coverage through machine learning. *Remote Sens. Environ.* **2019**, *222*, 286–302. [CrossRef]

63. Dreano, D.; Mallick, B.; Hoteit, I. Filtering remotely sensed chlorophyll concentrations in the Red Sea using a space–time covariance model and a Kalman filter. *Spat. Stat.* **2015**, *13*, 1–20. [CrossRef]

64. Jouini, M.; Levy, M.; Crépon, M.; Thiria, S. Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sens. Environ.* **2013**, *131*, 232–246. [CrossRef]

65. Krasnopolsky, V.; Nadiga, S.; Mehra, A.; Bayler, E.; Behringer, D. Neural networks technique for filling gaps in satellite measurements: Application to ocean color observations. *Comput. Intell. Neurosci.* **2016**, *2016*, 1–9. [CrossRef] [PubMed]

66. Liu, D.; Wang, Y. Trends of satellite derived chlorophyll-a (1997–2011) in the Bohai and Yellow Seas, China: Effects of bathymetry on seasonal and inter-annual patterns. *Prog. Oceanogr.* **2013**, *116*, 154–166. [CrossRef]

67. Park, J.; Kim, J.H.; Kim, H.C.; Kim, B.K.; Bae, D.; Jo, Y.H.; Jo, N.; Lee, S.H. Reconstruction of Ocean Color Data Using Machine Learning Techniques in Polar Regions: Focusing on Off Cape Hallett, Ross Sea. *Remote Sens.* **2019**, *11*, 1366. [CrossRef]

68. Saulquin, B.; Gohin, F.; d'Andon, O.F. Interpolated fields of satellite-derived multi-algorithm chlorophyll-a estimates at global and European scales in the frame of the European Copernicus-Marine Environment Monitoring Service. *J. Oper. Oceanogr.* **2019**, *12*, 47–57.

69. Shunmugapandi, R.; Inamdar, A.B.; Gedam, S.K. Long-time-scale investigation of phytoplankton communities based on their size in the Arabian Sea. *Int. J. Remote Sens.* **2020**, *41*, 5992–6009.

70. Han, Z.; He, Y.; Liu, G.; Perrie, W. Application of DINCAE to Reconstruct the Gaps in Chlorophyll-a Satellite Observations in the South China Sea and West Philippine Sea. *Remote Sens.* **2020**, *12*, 480. [CrossRef]

71. Jayaram, C.; Priyadarshi, N.; Kumar, J.P.; Bhaskar, T.V.S.U.; Raju, D.; Kochuparampil, A.J. Analysis of gap-free chlorophyll-a data from MODIS in Arabian Sea, reconstructed using DINEOF. *Int. J. Remote Sens.* **2018**, *39*, 7506–7522.

72. Jayaram, C.; Jonnakuti, P.; Udaya Bhaskar, T.; Bhavani, I.; Rao, T.; Nagamani, P. Reconstruction of Gap-Free OCM-2 Chlorophyll-a Concentration Using DINEOF. *J. Indian Soc. Remote Sens.* **2021**, *49*, 1419–1425. [CrossRef]

73. Mohamed Yussof, F.N. Reconstruction of chlorophyll-a data by using DINEOF approach in Sepanggar Bay, Malaysia. *Comput. Sci.* **2021**, *16*, 345–356.

74. Mohebzadeh, H.; Mokari, E.; Daggupati, P.; Biswas, A. A machine learning approach for spatiotemporal imputation of MODIS chlorophyll-a. *Int. J. Remote Sens.* **2021**, *42*, 7381–7404.

75. Park, J.; Kim, H.C.; Bae, D.; Jo, Y.H. Data Reconstruction for Remotely Sensed Chlorophyll-a Concentration in the Ross Sea Using Ensemble-Based Machine Learning. *Remote Sens.* **2020**, *12*, 11898. [CrossRef]

76. Pottier, C.; Garçon, V.; Larnicol, G.; Sudre, J.; Schaeffer, P.; Le Traon, P.-Y. Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic Using Weighted Averaging and Objective Analysis. *IEEE Trans. Geosci. Remote. Sens.* **2006**, *44*, 3436–3451. [CrossRef]

77. Volpe, G.; Nardelli, B.B.; Cipollini, P.; Santoleri, R.; Robinson, I.S. Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations. *Remote Sens. Environ.* **2012**, *117*, 223–235. [CrossRef]

78. Wang, Y.; Liu, D. Reconstruction of satellite chlorophyll-a data using a modified DINEOF method: A case study in the Bohai and Yellow seas, China. *Int. J. Remote Sens.* **2014**, *35*, 204–217. [CrossRef]

79. Liu, X.; Wang, M. Filling the Gaps of Missing Data in the Merged VIIRS SNPP/NOAA-20 Ocean Color Product Using the DINEOF Method. *Remote Sens.* **2019**, *11*, 178. [CrossRef]

80. Pottier, C.; Turiel, A.; Garçon, V. Inferring missing data in satellite chlorophyll maps using turbulent cascading. *Remote Sens. Environ.* **2008**, *112*, 4242–4260. [CrossRef]

81. Rebekah, S.; Inamdar, A.; Gedam, S.S. Long-Time-Scale Investigation of Phytoplankton Biomass Through Reconstructed Chlorophyll-A Data Using DINEOF Method. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7995–7998. [CrossRef]

82. Rinaldi, E.; Buongiorno Nardelli, B.; Volpe, G.; Santoleri, R. Chlorophyll distribution and variability in the Sicily Channel (Mediterranean Sea) as seen by remote sensing data. *Cont. Shelf Res.* **2014**, *77*, 61–68. [CrossRef]

83. Casey, B.; Arnone, R.; Flynn, P. Simple and efficient technique for spatial/temporal composite imagery. In *Proceedings of the Coastal Ocean Remote Sensing*; Frouin, R.J., Lee, Z., Eds.; International Society for Optics and Photonics—SPIE: Bellingham, WA, USA, 2007; Volume 6680, p. 668014. [CrossRef]

84. Kwiatkowska, E.; Fargion, G. Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data. *Geosci. Remote Sens. IEEE Trans.* **2004**, *41*, 2844–2860. [CrossRef]

85. Andreo, V.C.; Dogliotti, A.I.; Tauro, C.B. Remote Sensing of Phytoplankton Blooms in the Continental Shelf and Shelf-Break of Argentina: Spatio-Temporal Changes and Phenology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 5315–5324. [CrossRef]

86. Racault, M.-F.; Le Quéré, C.; Buitenhuis, E.; Sathyendranath, S.; Platt, T. Phytoplankton phenology in the global ocean. *Ecol. Indic.* **2012**, *14*, 152–163. [CrossRef]

87. McGinty, N.; Guðmundsson, K.; Ágústsdóttir, K.; Marteinsdóttir, G. Environmental and climactic effects of chlorophyll-a variability around Iceland using reconstructed satellite data fields. *J. Mar. Syst.* **2016**, *163*, 31–42. [CrossRef]

88. Li, Y.; He, R. 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.* **2014**, *144*, 98–108. [CrossRef]

89. Miles, T.N.; He, R.; Li, M. Characterizing the South Atlantic Bight seasonal variability and cold-water event in 2003 using a daily cloud-free SST and chlorophyll analysis. *Geophys. Res. Lett.* **2009**, *36*.

90. Shropshire, T.; Li, Y.; He, R. Storm impact on sea surface temperature and chlorophyll a in the Gulf of Mexico and Sargasso Sea based on daily cloud-free satellite data reconstructions. *Geophys. Res. Lett.* **2016**, *43*, 12199–12207.

91. Waite, J.N.; Mueter, F.J. 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.* **2013**, *116*, 179–192. [CrossRef]

92. Zhao, Y.; He, R. Cloud-free sea surface temperature and colour reconstruction for the Gulf of Mexico: 2003–2009. *Remote Sens. Lett.* **2012**, *3*, 697–706. [CrossRef]

93. Alvera-Azcárate, A.; Barth, A.; Beckers, J.M.; Weisberg, R.H. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *J. Geophys. Res. Ocean.* **2007**, *112* [CrossRef]

94. Barth, A.; Alvera-Azcárate, A.; Troupin, C.; Beckers, J.M. DINCAE 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. *Geosci. Model Dev.* **2022**, *15*, 2183–2196. [CrossRef]

95. Miles, T.N.; He, R. Temporal and spatial variability of Chl-a and SST on the South Atlantic Bight: Revisiting with cloud-free reconstructions of MODIS satellite imagery. *Cont. Shelf Res.* **2010**, *30*, 1951–1962. [CrossRef]

96. Chapman, C.; Charantonis, A.A. Reconstruction of Subsurface Velocities From Satellite Observations Using Iterative Self-Organizing Maps. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 617–620. [CrossRef]

97. Gregg, W.; Esaias, W.; Feldman, G.; Frouin, R.; Hooker, S.; McClain, C.; Woodward, R. Coverage opportunities for global ocean color in a multimission era. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1620–1627. [CrossRef]

98. Gregg, W.; Woodward, R. Improvements in coverage frequency of ocean color: Combining data from SeaWiFS and MODIS. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1350–1353. [CrossRef]

99. Preußer, A.; Willmes, S.; Heinemann, G.; Paul, S. Thin-ice dynamics and ice production in the Storfjorden polynya for winter seasons 2002/2003–2013/2014 using MODIS thermal infrared imagery. *Cryosphere* **2015**, *9*, 1063–1073. [CrossRef]

100. Fieguth, P.; Menemenlis, D.; Ho, T.; Willsky, A.; Wunsch, C. Mapping Mediterranean Altimeter Data with a Multiresolution Optimal Interpolation Algorithm. *J. Atmos. Ocean. Technol.* **1998**, *15*, 535–546. [CrossRef]

101. Hamlington, B.; Leben, R.; Strassburg, M.; Kim, K.Y. Cyclostationary empirical orthogonal function sea-level reconstruction. *Geosci. Data J.* **2014**, *1*, 13–19 [CrossRef]

102. Kumar, P.; Hamlington, B.; Cheon, S.H.; Han, W.; Thompson, P. 20th Century Multivariate Indian Ocean Regional Sea Level Reconstruction. *J. Geophys. Res. Ocean.* **2020**, *125*. [CrossRef]

103. Chambers, D.P.; Mehlhaff, C.A.; Urban, T.J.; Fujii, D.; Nerem, R.S. Low-frequency variations in global mean sea level: 1950–2000. *J. Geophys. Res. Ocean.* **2002**, *107*, 1-1–1-10. [CrossRef]

104. Hamlington, B.D.; Leben, R.R.; Wright, L.A.; Kim, K.Y. Regional Sea Level Reconstruction in the Pacific Ocean. *Mar. Geod.* **2012**, *35*, 98–117. [CrossRef]

105. Barth, A.; Alvera-Azcárate, A.; Licer, M.; Beckers, J.M. DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geosci. Model Dev.* **2020**, *13*, 1609–1622. [CrossRef]

106. Jung, S.; Yoo, C.; Im, J. High-Resolution Seamless Daily Sea Surface Temperature Based on Satellite Data Fusion and Machine Learning over Kuroshio Extension. *Remote Sens.* **2022**, *14*, 575. [CrossRef]

107. Park, S.; Chu, P. Interannual SST variability in the Japan/East Sea and relationship with environmental variables. *J. Oceanogr.* **2006**, *62*, 115–132. [CrossRef]

108. Reynolds, R.W.; Smith, T.M.; Liu, C.; Chelton, D.B.; Casey, K.S.; Schlax, M.G. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *J. Clim.* **2007**, *20*, 5473–5496. [CrossRef]

109. Reynolds, R.W.; Smith, T.M. Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *J. Clim.* **1994**, *7*, 929–948. [CrossRef]

110. Saulquin, B.; Gohin, F. Mean seasonal cycle and evolution of the sea surface temperature from satellite and in situ data in the English Channel for the period 1986–2006. *Int. J. Remote Sens.* **2010**, *31*, 4069–4093. [CrossRef]

111. She, J.; Høyer, J.L.; Larsen, J. Assessment of sea surface temperature observational networks in the Baltic Sea and North Sea. *J. Mar. Syst.* **2007**, *65*, 314–335. [CrossRef]

112. Bisagni, J.; Seemann, K.; Mavor, T. High-resolution satellite-derived sea-surface temperature variability over the Gulf of Maine and Georges Bank region, 1993–1996. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2001**, *48*, 71–94. [CrossRef]

113. Chao, Y.; Li, Z.; Farrara, J.D.; Hung, P. Blending Sea Surface Temperatures from Multiple Satellites and In Situ Observations for Coastal Oceans. *J. Atmos. Ocean. Technol.* **2009**, *26*, 1415–1426. [CrossRef]

114. Guan, L.; Kawamura, H. Merging Satellite Infrared and Microwave SSTs: Methodology and Evaluation of the New SST. *J. Oceanogr.* **2004**, *60*, 905–912. [CrossRef]

115. Hirahara, N.; Sonogashira, M.; Kasahara, H.; Iiyama, M. Denoising and Inpainting of Sea Surface Temperature Image with Adversarial Physical Model Loss. In Proceedings of the Pattern Recognition: 5th Asian Conference, ACPR 2019, Auckland, New Zealand, 26–29 November 2019; Revised Selected Papers, Part I; Springer: Berlin/Heidelberg, Germany, 2019; pp. 339–352. [CrossRef]

116. Houseago-Stokes, R.E.; Challenor, P.G. Using PPCA to Estimate EOFs in the Presence of Missing Values. *J. Atmos. Ocean. Technol.* **2004**, *21*, 1471–1480. [CrossRef]

117. Kawai, Y.; Kawamura, H.; Takahashi, S.; Hosoda, K.; Murakami, H.; Kachi, M.; Guan, L. Satellite-based high-resolution global optimum interpolation sea surface temperature data. *J. Geophys. Res. Ocean.* **2006**, *111*. [CrossRef]

118. Lguensat, R.; Tandeo, P.; Fablet, R.; Garello, R. Spatio-temporal interpolation of Sea Surface Temperature using high resolution remote sensing data. In Proceedings of the 2014 Oceans, St. John's, NL, Canada, 14–19 September 2014. [CrossRef]

119. Mauri, E.; Poulain, P.M.; Notarstefano, G. Spatial and temporal variability of the sea surface temperature in the Gulf of Trieste between January 2000 and December 2006. *J. Geophys. Res. Ocean.* **2008**, *113*. [CrossRef]

120. Ouala, S.; Fablet, R.; Herzet, C.; Chapron, B.; Pascual, A.; Collard, F.; Gaultier, L. Neural Network Based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-Derived Sea Surface Temperature. *Remote Sens.* **2018**, *10*, 1864. [CrossRef]

121. Pisoni, E.; Pastor, F.; Volta, M. Artificial Neural Networks to reconstruct incomplete satellite data: Application to the Mediterranean Sea Surface Temperature. *Nonlinear Process. Geophys.* **2008**, *15*, 61–70. [CrossRef]

122. Shibata, S.; Iiyama, M.; Hashimoto, A.; Minoh, M. Restoration of Sea Surface Temperature Satellite Images Using a Partially Occluded Training Set. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 2771–2776. [CrossRef]

123. Garcia-Gorriz, E.; Garcia-Sanchez, J. Prediction of sea surface temperatures in the western Mediterranean Sea by neural networks using satellite observations. *Geophys. Res. Lett.* **2007**, *34*. [CrossRef]

124. Kang, S.H.; Choi, Y.; Choi, J.Y. Restoration of Missing Patterns on Satellite Infrared Sea Surface Temperature Images Due to Cloud Coverage Using Deep Generative Inpainting Network. *J. Mar. Sci. Eng.* **2021**, *9*, 310. [CrossRef]

125. Ping, B.; Su, F.; Meng, Y. Reconstruction of Satellite-Derived Sea Surface Temperature Data Based on an Improved DINEOF Algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4181–4188. [CrossRef]

126. Yang, Y.; Dong, J.; Sun, X.; Lguensat, R.; Jian, M.; Wang, X. Ocean Front Detection From Instant Remote Sensing SST Images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1960–1964. [CrossRef]

127. Youzhuan, D.; Wei, Z.; Zhihua, M.; Xiaofei, W.; Delu, P. Reconstruction of incomplete satellite SST data sets based on EOF method. *Acta Oceanol. Sin.* **2009**, *28*, 36–44.

128. Alvera-Azcárate, A.; Barth, A.; Sirjacobs, D.; Beckers, J.M. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Sci.* **2009**, *5*, 475–485. [CrossRef]

129. Paul, S.; Willmes, S.; Gutjahr, O.; Preußer, A.; Heinemann, G. Spatial Feature Reconstruction of Cloud-Covered Areas in Daily MODIS Composites. *Remote Sens.* **2015**, *7*, 5042–5056. [CrossRef]

130. Nechad, B.; Alvera-Azcaràte, A.; Ruddick, K.; Greenwood, N. Reconstruction of MODIS total suspended matter time series maps by DINEOF and validation with autonomous platform data. *Ocean Dyn.* **2011**, *61*, 1205–1214. [CrossRef]

131. Conkright, M.E.; Gregg, W.W. Comparison of global chlorophyll climatologies: In situ, CZCS, Blended in situ -CZCS and SeaWiFS. *Int. J. Remote Sens.* **2003**, *24*, 969–991. [CrossRef]

132. Brody, S.R.; Lozier, M.S.; Dunne, J.P. A comparison of methods to determine phytoplankton bloom initiation. *J. Geophys. Res. Ocean.* **2013**, *118*, 2345–2357. [CrossRef]

133. Fu, Y.; Xu, S.; Zhang, C.; Sun, Y. Spatial downscaling of MODIS Chlorophyll-a using Landsat 8 images for complex coastal water monitoring. *Estuar. Coast. Shelf Sci.* **2018**, *209*, 149–159. [CrossRef]

134. Jo, Y.H.; Kim, D.w.; Kim, H. Chlorophyll concentration derived from microwave remote sensing measurements using artificial neural network algorithm. *J. Mar. Sci. Technol.* **2018**, *26*, 102–110. [CrossRef]

135. Hilborn, A.; Costa, M. Applications of DINEOF to Satellite-Derived Chlorophyll-a from a Productive Coastal Region. *Remote Sens.* **2018**, *10*, 1449. [CrossRef]

136. Maritorena, S.; Siegel, D.A. Consistent merging of satellite ocean color data sets using a bio-optical model. *Remote Sens. Environ.* **2005**, *94*, 429–440. [CrossRef]

137. Kwiatkowska, E.; Fargion, G. Merger of ocean color information from multiple satellite missions under the NASA SIMBIOS Project Office. In Proceedings of the Fifth International Conference on Information Fusion, FUSION (IEEE Cat.No.02EX5997), Annapolis, MD, USA, 8–11 July 2002; Volume 1, pp. 291–298. [CrossRef]

138. Navarro, G.; Caballero, I.; Prieto, L.; Vázquez, A.; Flecha, S.; Huertas, I.; Ruiz, J. Seasonal-to-interannual variability of chlorophyll-a bloom timing associated with physical forcing in the Gulf of Cádiz. *Adv. Space Res.* **2012**, *50*, 1164–1172. [CrossRef]

139. Michel, R.; Beckers, J.M.; Alvarez, A.; Tintoré, J. Results on SSH neural network forecasting in the Mediterranean Sea. In *Remote Sensing of the Ocean and Sea Ice 2001*; SPIE: Bellingham, WA, USA, 2002; Volume 4544. [CrossRef]

140. Høyer, J.L.; She, J. Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea. *J. Mar. Syst.* **2007**, *65*, 176–189. [CrossRef]

141. Lguensat, R.; Tandeo, P.; Ailliot, P.; Chapron, B.; Fablet, R. Using archived datasets for missing data interpolation in ocean remote sensing observation series. In Proceedings of the OCEANS 2016, Shanghai, China, 10–13 April 2016; pp. 1–5. [CrossRef]

142. Ping, B.; Su, F.; Meng, Y. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLoS ONE* **2016**, *11*, e0155928. [CrossRef]

143. Huynh, H.N. Reconstruction and analysis of long-term satellite-derived sea surface temperature for the South China Sea. *J. Oceanogr.* **2016**, *72.* [CrossRef]

144. Nikolaidis, A.; Georgiou, G.; Hadjimitsis, D.; Akylas, E. Applying DINEOF Algorithm On Cloudy Sea-Surface Temperature Satellite Data Over The Eastern Mediterranean Sea. *First Int. Conf. Remote. Sens. Geoinf. Environ.* **2013**, *8795*, 176–185. [CrossRef]

145. Tandeo, P.; Autret, E.; Chapron, B.; Fablet, R.; Garello, R. SST spatial anisotropic covariances from METOP-AVHRR data. *Remote. Sens. Environ.* **2014**, *141*, 144–148. [CrossRef]

146. Thiébaux, J.; Rogers, E.; Wang, W.; Katz, B. A New High-Resolution Blended Real-Time Global Sea Surface Temperature Analysis. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 645–656. [CrossRef]

147. Patil, K.; Deo, M. Prediction of daily sea surface temperature using efficient neural networks. *Ocean Dyn.* **2017**, *67*, 357–368 [CrossRef]

148. Liu, X.; Wang, M. Global daily gap-free ocean color products from multi-satellite measurements. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102714. [CrossRef]

149. Wang, M.; Jiang, L.; Son, S.; Liu, X.; Voss, K.J. Deriving consistent ocean biological and biogeochemical products from multiple satellite ocean color sensors. *Opt. Express* **2020**, *28*, 2661–2682. [CrossRef]

150. Álvarez, A.; López, C.; Riera, M.; Hernández-García, E.; Tintoré, J. Forecasting the SST Space-time variability of the Alboran Sea with genetic algorithms. *Geophys. Res. Lett.* **2000**, *27*, 2709–2712. [CrossRef]

151. Beckers, J.M.; Rixen, M. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *J. Atmos. Ocean. Technol.* **2003**, *20*, 1839–1856. [CrossRef]

152. Chin, T.M.; Vazquez-Cuervo, J.; Armstrong, E.M. A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.* **2017**, *200*, 154–169. [CrossRef]

153. Everson, R.; Cornillon, P.; Sirovich, L.; Webber, A. An Empirical Eigenfunction Analysis of Sea Surface Temperatures in the Western North Atlantic. *J. Phys. Oceanogr.* **1997**, *27*, 468–479. [CrossRef]

154. He, R.; Weisberg, R.H.; Zhang, H.; Muller-Karger, F.E.; Helber, R.W. A cloud-free, satellite-derived, sea surface temperature analysis for the West Florida Shelf. *Geophys. Res. Lett.* **2003**, *30*, 4-1–4-5

155. Mohebzadeh, H.; Lee, T. Spatial downscaling of MODIS Chlorophyll-a with machine learning techniques over the west coast of the Yellow Sea in South Korea. *J. Oceanogr.* **2020**, *77.* [CrossRef]

156. Hu, C.; Feng, L.; Lee, Z.; Franz, B.; Bailey, S.; Werdell, J.; Proctor, C. Improving Satellite Global Chlorophyll a Data Products Through Algorithm Refinement and Data Recovery. *J. Geophys. Res. Ocean.* **2019**, *124.* [CrossRef]

157. Stock, A.; Subramaniam, A.; Van Dijken, G.L.; Wedding, L.M.; Arrigo, K.R.; Mills, M.M.; Cameron, M.A.; Micheli, F. Comparison of Cloud-Filling Algorithms for Marine Satellite Data. *Remote Sens.* **2020**, *12*, 3313. [CrossRef]

158. Schaeffer, P.; Faugére, Y.; Legeais, J.F.; Ollivier, A.; Guinle, T.; Picot, N. The CNES_CLS11 Global Mean Sea Surface Computed from 16 Years of Satellite Altimeter Data. *Mar. Geod.* **2012**, *35*, 3–19. [CrossRef]

159. Saxena, N. Efficient Downscaling of Satellite Oceanographic Data With Convolutional Neural Networks. In Proceedings of the 28th International Conference on Advances in Geographic Information Systems; Association for Computing Machinery, SIGSPATIAL '20, New York, NY, USA, 3–6 November 2020; pp. 659–660. [CrossRef]

160. Liu, J.; Sun, Y.; Ren, K.; Zhao, Y.; Deng, K.; Wang, L. A Spatial Downscaling Approach for WindSat Satellite Sea Surface Wind Based on Generative Adversarial Networks and Dual Learning Scheme. *Remote Sens.* **2022**, *14*, 769. [CrossRef]

161. Atkinson, P.M. Downscaling in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *22*, 106–114. [CrossRef]

162. Fablet, R.; Huynh Viet, P.; Lguensat, R.; Horrein, P.H.; Chapron, B. Spatio-Temporal Interpolation of Cloudy SST Fields Using Conditional Analog Data Assimilation. *Remote Sens.* **2018**, *10*, 310. [CrossRef]

163. Kalinić, H.; Bilokapić, Z.; Matić, F. Can Local Geographically Restricted Measurements Be Used to Recover Missing Geo-Spatial Data? *Sensors* **2021**, *21*, 3507. [CrossRef]

164. Kalinić, H.; Ćatipović, L.; Matić, F. Optimal Sensor Placement Using Learning Models; A Mediterranean Case Study. *Remote Sens.* **2022**, *14*, 2989. [CrossRef]

165. Ocean Optics Web Book. Available online: https://www.oceanopticsbook.info/ (accessed on 27 December 2022).

166. O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res. Ocean.* **1998**, *103*, 24937–24953. [CrossRef]

167. O'Reilly, J. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. *SeaWiFS Postlaunch Calibration Valid. Anal.* **2000**, *11*, 9–23.

168. Hu, C.; Lee, Z.; Franz, B. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *J. Geophys. Res.* **2012**, *117*, C01011. [CrossRef]

169. Brewin, B.; Raitsos, D.; Dall'Olmo, G.; Zarokanellos, N.; Jackson, T.; Racault, M.F.; Boss, E.; Sathyendranath, S.; Jones, B.; Hoteit, I. Regional ocean-colour chlorophyll algorithms for the Red Sea. *Remote Sens. Environ.* **2015**, *165*, 64–85. [CrossRef]

170. Darecki, M.; Stramski, D. An evaluation of MODIS and Sea WiFS bio-optical algorithms in the Baltic sea. *Remote Sens. Environ.* **2004**, *89*, 326–350. [CrossRef]

171. Garcia, C.; Garcia, V.; Mcclain, C. Evaluation of SeaWiFS chlorophyll algorithms in the Southwestern Atlantic and Southern Oceans. *Remote Sens. Environ.* **2005**, *95*, 125–137. [CrossRef]

172. Claustre, H.; Morel, A.; Hooker, S.B.; Babin, M.; Antoine, D.; Oubelkheir, K.; Bricaud, A.; Leblanc, K.; Quéguiner, B.; Maritorena, S. Is desert dust making oligotrophic waters greener? *Geophys. Res. Lett.* **2002**, *29*, 107-1–107-4.

173. D'Ortenzio, F.; Santoleri, R.; Marullo, S.; Ragni, M.; Ribera d'Alcala, M. Empirical SeaWiFS chlorophyll algorithm validation for the Mediterranean Sea. In *Remote Sensing of the Ocean and Sea Ice 2000*; SPIE: Bellingham, WA, USA, 2000. [CrossRef]

174. Volpe, G.; Santoleri, R.; Vellucci, V.; Ribera d'Alcalà, M.; Marullo, S.; D'Ortenzio, F. The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sens. Environ.* **2007**, *107*, 625–638. [CrossRef]

175. Minnett, P.; Barton, I. Chapter 6 Remote Sensing of the Earth's Surface Temperature. *Exp. Methods Phys. Sci.* **2010**, *43*, 333–391. [CrossRef]

176. Stewart, R.H. *Methods of Satellite Oceanography*; University of California Press: Berkeley, CA, USA, 1985.

177. Wick, G.A. Infrared and microwave remote sensing of sea surface temperature. In *Seminar at the University of Colorado at Boulder "Remote Sensing Seminar" Graduate, 11th Conference on Interaction of the Sea and Atmosphere, Session 7.14*; AMS: Boston, MA, USA, 2002.

178. Maurer, J. Infrared and Microwave Remote Sensing of Sea Surface Temperature (SST). Available online: http://www2.hawaii.edu/jmaurer/sst/ (accessed on 2 February 2023).

179. Allison, L.; Kennedy, J.; Aeronautics, U.S.N.; Administration, S.; Center, G.S.F. *An Evaluation of Sea Surface Temperature as Measured by the Nimbus I High Resolution Infrared Radiometer*; NASA Technical Note; National Aeronautics and Space Administration: Washington, DC, USA, 1967.

180. McMillin, L.M. Estimation of sea surface temperatures from two infrared window measurements with different absorption. *J. Geophys. Res.* **1975**, *80*, 5113–5117.

181. Anding, D.; Kauth, R. Estimation of sea surface temperature from space. *Remote Sens. Environ.* **1970**, *1*, 217–220. [CrossRef]

182. Prabhakara, C.; Dalu, G.; Kunde, V.G. Estimation of sea surface temperature from remote sensing in the 11- to 13-m window region. *J. Geophys. Res.* **1974**, *79*, 5039–5044.

183. Walton, C.C.; Pichel, W.G.; Sapper, J.F.; May, D.A. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *J. Geophys. Res. Ocean.* **1998**, *103*, 27999–28012. [CrossRef]

184. Llewellyn-Jones, D.T.; Minnett, P.J.; Saunders, R.W.; Zavody, A.M. Satellite multichannel infrared measurements of sea surface temperature of the N.E. Atlantic Ocean using AVHRR/2. *Q. J. R. Meteorol. Soc.* **1984**, *110*, 613–631.

185. Deschamps, P.Y.; Phulpin, T. Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 mm. *Bound.-Layer Meteorol.* **1980**, *18*, 131–143. [CrossRef]

186. Kilpatrick, K.A.; Podestá, G.P.; Evans, R. Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res. Ocean.* **2001**, *106*, 9179–9197. [https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999JC000065]. [CrossRef]

187. Brown, O.B.; Minnett, P.J.; Evans, R.; Kearns, E.; Kilpatrick, K.; Kumar, A.; Sikorski, R.; Závody, A. MODIS infrared sea surface temperature algorithm algorithm theoretical basis document version 2.0. *Univ. Miami* **1999**, *31*, 098-33.
188. McClain, E.; Pichel, W.; Walton, C.; Ahmad, Z.; Sutton, J. Multi-channel improvements to satellite-derived global sea surface temperatures. *Adv. Space Res.* **1983**, *2*, 43–47. [CrossRef]
189. Schluessel, P.; Shin, H.Y.; Emery, W.J.; Grassl, H. Comparison of satellite-derived sea surface temperatures with in situ skin measurements. *J. Geophys. Res. Ocean.* **1987**, *92*, 2859–2874.
190. NASA Ocean Color–Algorithm Descriptions. Available online: https://oceancolor.gsfc.nasa.gov/atbd/ (accessed on 27 December 2022).
191. Beckers, J.M.; Barth, A.; Alvera-Azcárate, A. DINEOF reconstruction of clouded images including error maps. Application to the Sea-Surface Temperature around Corsican Island. *Ocean Sci. Discuss. OSD* **2006**, *2*. [CrossRef]
192. Dong, J.; Yin, R.; Sun, X.; Li, Q.; Yang, Y.; Qin, X. Inpainting of Remote Sensing SST Images With Deep Convolutional Generative Adversarial Network. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 173–177. [CrossRef]
193. Alvera-Azcárate, A.; Barth, A.; Rixen, M.; Beckers, J. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. *Ocean Model.* **2005**, *9*, 325–346. [CrossRef]
194. Sorjamaa, A.; Lendasse, A.; Yves, C.; Deleersnijder, E. An improved methodology for filling missing values in spatiotemporal climate data set. *Comput. Geosci.* **2010**, *14*, 55–64. [CrossRef]
195. Azcarate, A.; Barth, A.; Sirjacobs, D.; Lenartz, F.; Beckers, J.M. Data Interpolating Empirical Orthogonal Functions (DINEOF): A tool for geophysical data analyses. *Mediter. Mar. Sci.* **2011**, *12*. [CrossRef]
196. Lawson, C.L.; Hanson, R.J. Solving least squares problems. In *Classics in Applied Mathematics*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1976.
197. Preisendorfer, R.W.; Mobley, C.D. *Principal Component Analysis in Meteorology and Oceanography*; Elsevier: Amsterdam, The Netherlands, 1988.
198. Müller, K.R.; Smola, A.J.; Rätsch, G.; Schölkopf, B.; Kohlmorgen, J.; Vapnik, V. Predicting time series with support vector machines. In Proceedings of the Artificial Neural Networks—ICANN'97; Gerstner, W., Germond, A., Hasler, M., Nicoud, J.D., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; pp. 999–1004.
199. Boser, B.; Guyon, I.; Vapnik, V. A Training Algorithm for Optimal Margin Classifier. *Proc. Fifth Annu. ACM Workshop Comput. Learn. Theory* **1996**, *5*. [CrossRef]
200. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
201. Smola, A.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
202. Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2013.
203. Bernhard, S.; Smola, A.J. *Learning with kernels: Support Vector Machines, Regularization, Optimization and Beyond*; MIT Press: Cambridge, MA, USA, 2009.
204. Schölkopf, B.; Smola, A.; Smola, A.; Smola, A. Support Vector Machines and Kernel Algorithms. *Encycl. Biostat.* **2002**, *2005*, 5328–5335.
205. Bretherton, F.P.; Davis, R.E.; Fandry, C. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Res. Oceanogr. Abstr.* **1976**, *23*, 559–582. [CrossRef]
206. Virdee, T.S.; Kottegoda, N.T. A brief review of kriging and its application to optimal interpolation and observation well selection. *Hydrol. Sci. J.* **1984**, *29*, 367–387.
207. Burgess, T.M.; Webster, R. Optimal interpolation and Isarithmic mapping of Soil Properties. *J. Soil Sci.* **1980**, *31*, 315–331. [CrossRef]
208. Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46*, 175–185.
209. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
210. Tang, A.; Foong, J. A Qualitative Evaluation of Random Forest Feature Learning. In *Recent Advances on Soft Computing and Data Mining: Proceedings of The First International Conference on Soft Computing and Data Mining (SCDM-2014) Universiti Tun Hussein Onn Malaysia, Johor, Malaysia,16–18 June 2014*; Springer: Berlin/Heidelberg, Germany; Volume 287. [CrossRef]
211. Yu, P.S.; Yang, T.C.; Chen, S.Y.; Kuo, C.M.; Tseng, H.W. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J. Hydrol.* **2017**, *552*, 92–104. [CrossRef]
212. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. BF00337288. [CrossRef]
213. Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef]
214. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
215. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 27 December 2022).
216. Bebis, G.; Georgiopoulos, M. Feed-forward neural networks. *IEEE Potentials* **1994**, *13*, 27–31. [CrossRef]
217. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
218. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2014**, *63*, 139–144. [CrossRef]

219. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 1096–1103. [CrossRef]

220. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507.