Article

Application of Synthetic DINCAE–BME Spatiotemporal Interpolation Framework to Reconstruct Chlorophyll–a from Satellite Observations in the Arabian Sea

Xiting Yan [1], Zekun Gao [1], Yutong Jiang [1], Junyu He [1,2,3,*], Junjie Yin [1] and Jiaping Wu [1]

[1] Ocean College, Zhejiang University, Zhoushan 316000, China
[2] Ocean Academy, Zhejiang University, Zhoushan 316000, China
[3] Donghai Laboratory, Zhoushan 316000, China
* Correspondence: jyhe@zju.edu.cn

Abstract: Chlorophyll–a (Chl–a) concentration is an indicator of phytoplankton pigment, which is associated with the health of marine ecosystems. A commonly used method for the determination of Chl–a is satellite remote sensing. However, due to cloud cover, sun glint and other issues, remote sensing data for Chl–a are always missing in large areas. We reconstructed the Chl–a data from MODIS and VIIRS in the Arabian Sea within the geographical range of 12–28° N and 56–76° E from 2020 to 2021 by combining the Data Interpolating Convolutional Auto–Encoder (DINCAE) and the Bayesian Maximum Entropy (BME) methods, which we named the DINCAE–BME framework. The hold–out validation method was used to assess the DINCAE–BME method's performance. The root–mean–square–error (RMSE) and the mean–absolute–error (MAE) values for the hold–out cross–validation result obtained by the DINCAE–BME were 1.8824 mg m^{-3} and 0.4682 mg m^{-3}, respectively; compared with in situ Chl–a data, the RMSE and MAE values for the DINCAE–BME–generated Chl–a product were 0.6196 mg m^{-3} and 0.3461 mg m^{-3}, respectively. Moreover, DINCAE–BME exhibited better performance than the DINEOF and DINCAE methods. The spatial distribution of the Chl–a product showed that Chl–a values in the coastal region were the highest and the Chl–a values in the deep–sea regions were stable, while the Chl–a values in February and March were higher than in other months. Lastly, this study demonstrated the feasibility of combining the BME method and DINCAE.

Keywords: Chl–a; BME; DINCAE; DINEOF; reconstruction; machine learning

1. Introduction

As an indispensable part of marine ecology, phytoplankton have a vital impact on the ecology of marine organisms. Chlorophyll–a (Chl–a), as one of the major components of phytoplankton, is a valid biological indicator for marine primary productivity in aquatic environments [1,2]. Consequently, the study of Chl–a has great significance for marine ecosystems.

Commonly used methods for the determination of Chl–a include fluorescence extraction, spectrophotometry, high–performance liquid chromatography, satellite remote sensing and so on. Spectrophotometry is the most widely used method for measuring Chl–a, but the operation process is complicated and the extraction time is long [3]. High–performance liquid chromatography is a fast form of analysis but generally cannot be used for rapid analysis of a large number of samples in the field due to the numerous analysis steps [4]. The satellite remote sensing method is comprehensive, convenient and suitable for long–term dynamic monitoring of large–scale water bodies [5]. However, clouds, sun glint, aerosol and other problems always affect Chl–a monitoring carried out through satellite [6]. Therefore, many studies have focused on solving these problems by filling in

the missing data, proposing methods such as linear interpolations and complicated statistical methods [7,8]. Among them, the Data Interpolating Empirical Orthogonal Functions (DINEOF) method, based on an orthogonal empirical function and the correlation method from geostatistics, has been widely used in Chl–a interpolation [9–12].

DINEOF can make full use of truncated empirical orthogonal function (EOF) analysis, which is widely used in Chl–a reconstruction, to reconstruct missing data. It has the advantages of being able to minimizing iterative processing errors and conduct parameter–less processing with fast speed. DINEOF had been applied to interpolate missing data for Chl–a with good results in some regions, such as the Salish Sea [9], the Arabian Sea [10] and the Bohai Sea [11]. Furthermore, Ji et al. [12] used this method to reconstruct Chl–a data and sea surface temperature (SST) in order to evaluate the relationship between the two variables. The disadvantage of DINEOF is the truncated EOF series. Although a limited number of EOFs can be used to extract the key information from data, small–scale features not included in the most dominant EOFs may be lost. Convolution neural networks have the advantage of being able to capture small–scale features, and one representative method is Data Interpolating Convolutional Auto–Encoder (DINCAE). Compared with the DINEOF method, the DINCAE method based on a convolutional autoencoder structure has better computational efficiency and can retain more small–scale features, as well as demonstrating fewer reconstruction result errors and higher reconstruction accuracy [13]. Furthermore, overcoming the problem affecting DINEOF, which cannot handle nonlinear relationships in temporal and spatial domains, DINCAE can utilize the powerful capacity of a neural network to handle complex interactions and nonlinear relationships. In recent years, the DINCAE method has been widely used for reconstruction of missing data in satellite remote sensing data and has proven to be a potential tool in data reconstruction for SST [13–15], Chl–a [6,16,17] and other variables [18].

Geostatistical interpolation methods are also widely applied in filling in missing data. The Kriging method based on geostatistical interpolation has been gradually applied in research on marine science. Elena et al. [19] used the ordinary Kriging (OK) method as an interpolator to fill in missing satellite data, including for Chl–a, the surface partial pressure of carbon dioxide and particulate organic carbon, for the territorial waters of the province of Iraklion around the eastern Mediterranean island of Crete. Hou et al. [20] used the OK method to analyze the temporal and spatial evolution of Chl–a concentration in Dianchi Lake over the past 20 years. However, the Kriging method cannot process data with non–Gaussian distributions, make nonlinear predictions, deal with high–order attribute moments or utilize various types of uncertain and specific site information and core knowledge sources for interpolation purposes [21]. Overcoming the above deficiencies, Bayesian maximum entropy (BME) is the most advanced spatiotemporal interpolation method in geostatistics [22]. It combines the principle of maximum entropy and Bayes' theorem, which can integrate information content from different sources [23], and objectively considers the uncertainty of data. He et al. [21] improved the coverage of remote sensing Chl–a products by employing SST as an auxiliary variable. Jiang et al. [23] combined the BME method with DINEOF to interpolate XCO_2 data. In SST interpolation, Gao et al. [24] demonstrated that the BME method outperformed the optimum interpolation method and the OK method. He et al. [25] improved the accuracy of the BME method by utilizing a novel covariance method (i.e., the contigogram) in BME prediction. Lang et al. [26] combined the BME method with the physical oceanography formula to improve ocean pollution predictions.

Although it can achieve high–precision interpolation results by effectively integrating various types of information, the BME method is also highly limited by the distribution of data and the coverage percentage of the data. In the actual research process, the distribution of valid data in space and time may be uneven and involve missing data, which further limits the effectiveness of BME. Moreover, methods such as DINEOF and DINCAE only operate according to the input data, such as single–satellite data, without considering other sources of data and knowledge, which makes it difficult to guarantee the accuracy of the

interpolation results. The DINCAE method can be regarded as a method that generates a rudimentary result with a high data coverage percentage. Then, the output of the DINCAE method is input into the BME as soft data, which has to fine–tune the interpolation results by utilizing multi–source data. Use of the results from DINCAE as complementarity data can address the defects of BME, and BME inherits the DINCAE results, further improving the accuracy and effectiveness of DINCAE. In addition, the coverage rate for data from a single satellite is low, so it is not conducive to improving the accuracy of interpolation results if only the data from a single satellite are used for interpolation. Therefore, it is necessary to include several satellites to generate a precise Chl–a product with high spatiotemporal coverage.

The core contribution of this work is its proposal of the DINCAE–BME framework, which combines the principles of deep learning and spatiotemporal geostatistics and aims to improve the coverage of remote sensing data by using sea surface Chl–a products from the Aqua, Terra and S–NPP satellites. More importantly, this work proposes a general interpolation framework that can be applied in many other interpolation fields in addition to Chl–a interpolation when the auxiliary variables are unavailable, providing an alternative solution for research in other fields.

This work is organized as follows: the study area, data sources and methods used in this work are described in Section 2. Section 3 shows the comparison of the results from DINCAE–BME and other methods, along with a systematic analysis of the reconstruction results from DINCAE–BME. Section 4 discusses the methodological framework and Section 5 presents the conclusions of this work.

2. Materials and Methods

2.1. Study Area

The North Arabian Sea (NAS) (latitude 12–28° N and longitude 56–76° E), one of the most productive sea regions [27–29], was chosen as the study area. The NAS is located between the Horn of Africa (on the Somali Peninsula) and southern Asia (the Arabian Peninsula and Indian Peninsula). As marked by the red square in Figure 1, the NAS includes the Gulf of Oman and is adjacent to India, Pakistan, Iran and Oman. The NAS is in a tropical monsoon climate zone, with high temperatures all year round. The northeast monsoon prevails in autumn and winter, and precipitation is scarce. The southwest monsoon prevails in spring and summer, and the amount of precipitation is greater than in the other seasons. The yellow dots represent the locations of the Argo data, which were used as in situ Chl–a measurements in this work. Taking into account the timeliness of the data, the study period was two years from 2020 to 2021. Another reason for choosing this study period was that COVID–19 appeared in 2020 and spread around the world in 2021. This work further discusses the impact of COVID–19 on the marine environment by comparing the Chl–a interpolation results in 2020 and 2021.

2.2. Satellite Data

In this study, we used daily Chl–a data from VIIRS onboard the S–NPP satellite and MODIS onboard the Terra satellite and Aqua satellite, which were downloaded from https://oceancolor.gsfc.nasa.gov/l3/order/ accessed on 30 March 2022 at a spatial resolution of 4 km and a temporal resolution of 1 day. Due to cloud cover, sun glint and other issues, missing data in the NAS is extremely serious, especially in summer. The rate of missing data for Aqua over 2020–2021 for every pixel was calculated and is shown in Figure 2, illustrating that the missing data rate is close to 100% in some regions. During a two–year period of 731 days, the NAS was completely uncovered by remote sensing data for 90 days. It can be seen that the overall missing data rate was still relatively high, basically above 50%, and the missing data rate for Aqua was only 17%, which may have been mainly caused by clouds [10]. The monthly coverage curves for remote sensing data from Aqua, Terra and S–NPP from 2020 to 2021 are shown in Figure 3.
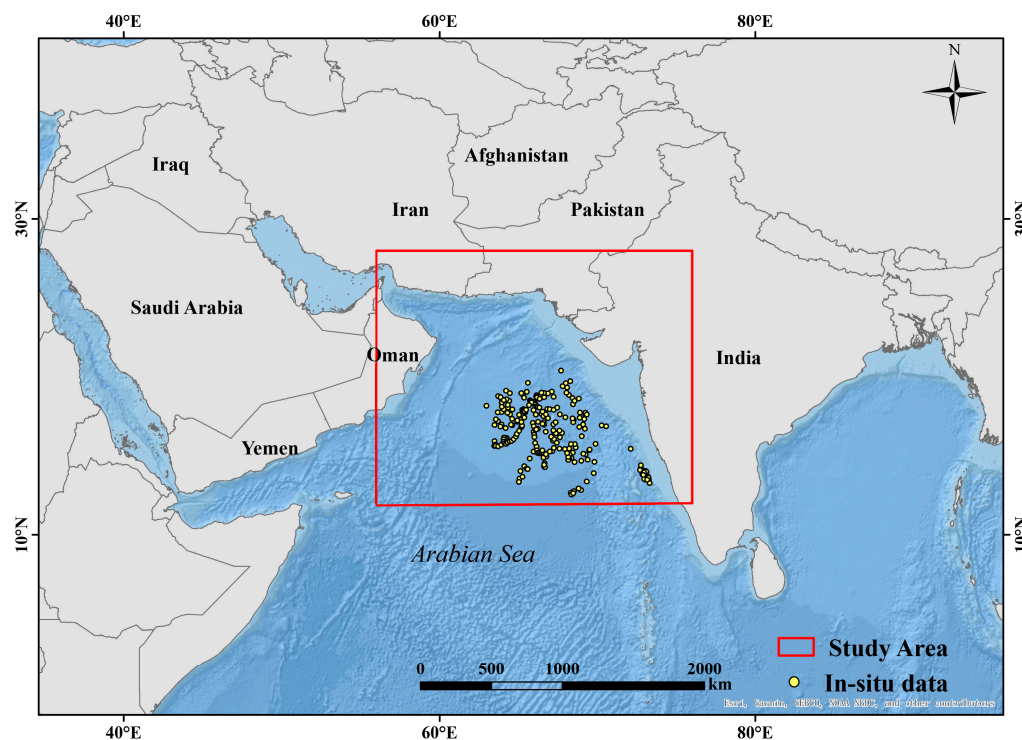
Figure 1. The red square represents the location of the NAS in the map of Asia. Yellow dots represent the locations of the in situ Chl–a measurements from the Argo data.
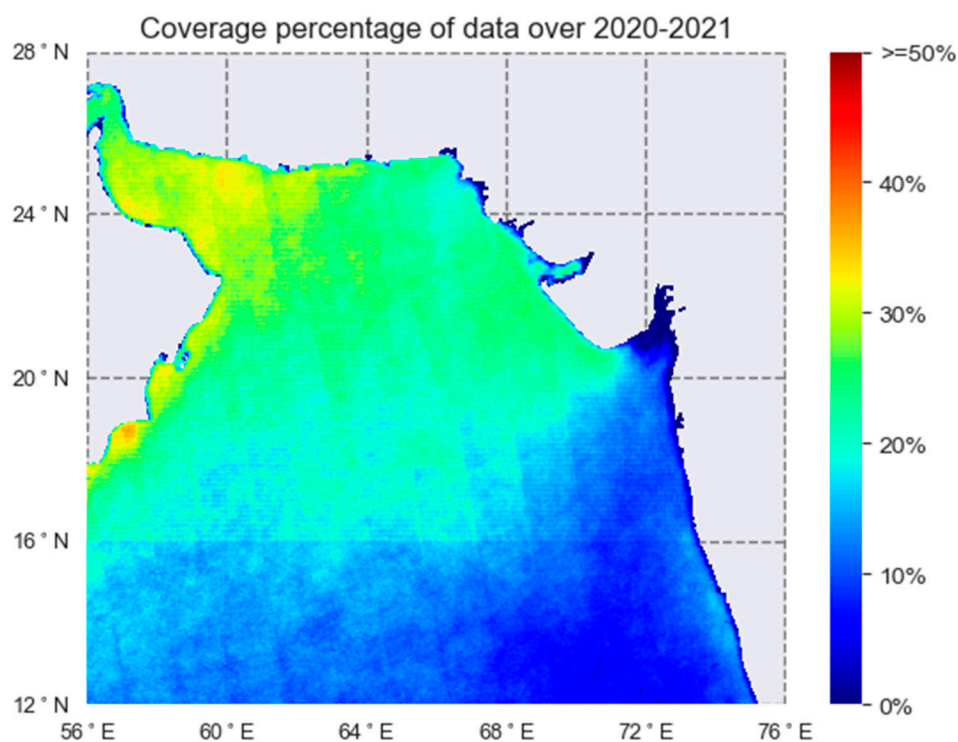

Figure 2. Percentages of Chl–a data coverage for the NAS from 2020 to 2021. The gray area represents the land.
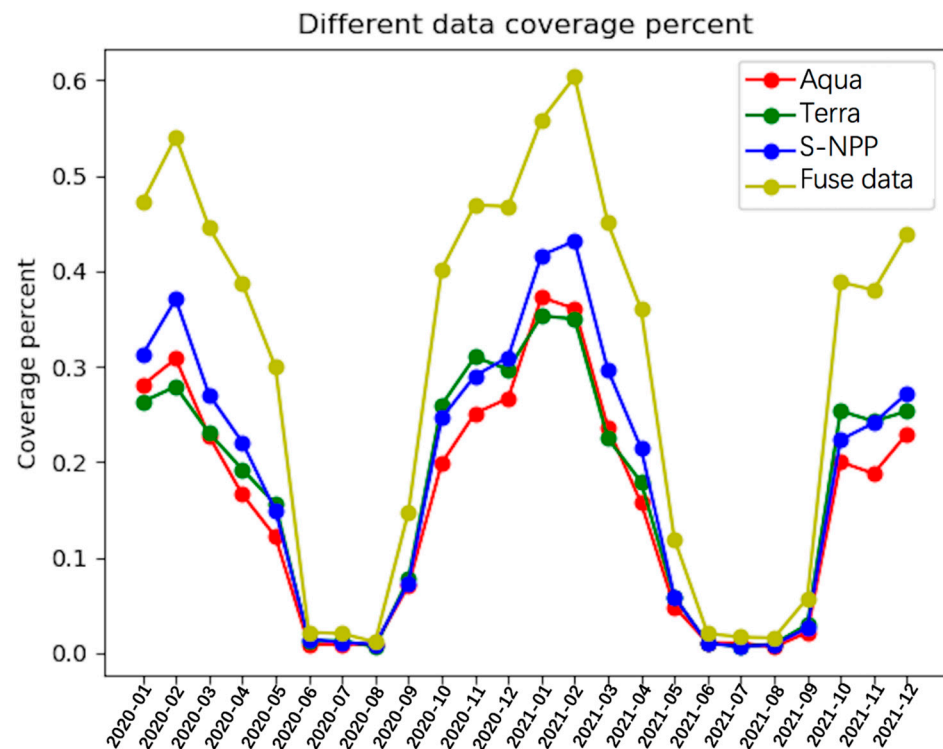
Figure 3. Aqua, Terra, S–NPP and the three satellites' combined monthly percentages of coverage from 2020 to 2021.

2.3. In Situ Data

The in situ Chl–a data were downloaded from https://dataselection.euro--argo.eu/ accessed on 30 March 2022 and obtained from the Argo buoys, which is an international program for measuring water properties across the world's oceans. The data provided by Argo undergo strict quality inspection and are updated in real time. In the research, 242 in situ Chl–a data values were matched with the satellite remote sensing Chl–a products during 2020–2021, and the values for the in situ Chl–a data ranged from 0.0146 to 4.7888 mg m^{-3}. The in situ data were used to validate the performance of the DINEOF, DINCAE and DINCAE–BME methods. The in situ data were matched with the nearest remote sensing data, including Aqua, Terra and S–NPP. The root–mean–square error (RMSE), the mean absolute error (MAE) and R^2 were calculated for each matched group of data. The results are shown in Table 1.

Table 1. Results were obtained by comparing Aqua, Terra and S–NPP data with in situ data.

Satellite	Count	RMSE (mg m^{-3})	MAE (mg m^{-3})	R^2
Aqua	59	0.3855	0.1255	0.7432
Terra	73	0.4237	0.1556	0.7246
S–NPP	67	0.4637	0.2368	0.6240

2.4. Cross–Validation

Given that the Aqua data had the highest precision in remote Chl–a estimation, they were regarded as the reference data in the current study. A variant of the hold–out method was used in this study to evaluate the performance of the methods. Therefore, the Aqua data were randomly split into three parts denoted as the "train data set" (80% of the data), "validation data set" (10% of the data) and "test data set" (10% of the data). Specifically, the train, validation and test datasets contained 10,426,064, 1,303,258 and 1,304,569 data values, respectively. The train dataset was used for DINEOF and DINCAE modeling and as the

hard data in the DINCAE–BME framework. The test dataset was used as reference data for generating soft data with uncertainty for BME modeling.

2.5. Methods

2.5.1. DINEOF

In the DINEOF method, the origin data matrix can be defined as X with S $\times$ N dimensions, where the spatial dimension of X is S and the temporal dimension of X is N. The process of singular value decomposition (SVD) can be described with the following equation:

$$X = UDV^T = \sum_{m=1}^{K} \rho_m U_m V_m^T \tag{1}$$

In this equation, U denotes the spatial decomposition vector, which has S $\times$ R dimensions, and V denotes the temporal decomposition vector, which has N $\times$ R dimensions. D is a diagonal matrix consisting of eigenvalues of dimensions R $\times$ R. ρ is the singular value and K is the number of singular values. Hence, when giving the specific U and V matrix, X can be reconstructed using Equation (1). However, the process of reconstruction with the given U and V matrix only works when the observation data are complete. In reality, observation data have large amounts of missing data that must be pre–processed before construction.

The EOF decomposition of a large amount of data is not affected by local minority changes, which means the reconstruction values for the missing area can be obtained by iteration. The process of iteration can be described as follows: first of all, missing data are filled with zeros and the missing regions are defined as I. The temporal average values at each location are removed from the data, resulting in a matrix defined as X_0. Then, the U, V and D matrix is obtained through SVD with X_0. The missing data for point $(i, j) \in$ I can be obtained by truncated EOF reconstruction according to the following equation:

$$(X)_{ij} = \left(U_T D_T (V_T)^T \right)_{ij} = \sum_{k=1}^{N} \rho_k (U_k)_i \left(V_k^T \right)_j \tag{2}$$

Iteration of the above process can be maintained until convergence. The expected maximum number of modes was 100, and the convergence threshold was set to 0.001. The code for DINEOF is available at https://github.com/aida--alvera/DINEOF accessed on 30 March 2022 and the work environment was Ubuntu 18.04.

2.5.2. DINCAE

Barth et al. [13] proposed DINCAE to reconstruct missing observations using training with incomplete satellite observations. Different from DINEOF, which is based on empirical orthogonal function analysis, DINCAE utilizes a convolutional neural network that has strong power in modeling correlations between spatial data and temporal data, allowing for the reconstruction of the missing data. The data preprocessing phase, the training phase and the reconstruction phase can all be grouped together to form the DINCAE reconstruction process.

In the process of data preprocessing, the time average of the Chl–a was subtracted to obtain the mean normalized result, and the missing data were set as zero. For the purpose of modeling the relativity between the spatial data and Chl–a data better, the longitude and latitude were scaled linearly between −1 and 1 using the following equation:

$$x' = 2 * \frac{x - x_{min}}{x_{max} - x_{min}} - 1 \tag{3}$$

The time data also had to be scaled through sine and cosine transforms in order to model the relativity between the temporal data and Chl–a data, for which the following equations were employed:

$$time_{sin} = \sin\left(\frac{2\pi * time}{365.25}\right) \tag{4}$$

$$time_{cos} = \cos\left(\frac{2\pi * time}{365.25}\right) \tag{5}$$

In conclusion, the input data included six variables, which were Chl–a, scaled longitude and latitude, sine and cosine transforms for time data and error variance.

During the training phase, the DINCAE neural network was trained iteratively by using the input data. As shown in Figure 4, the structure of DINCAE included five parts: an input layer, an encoder module, fully connected layers, a decoder module and an output layer.
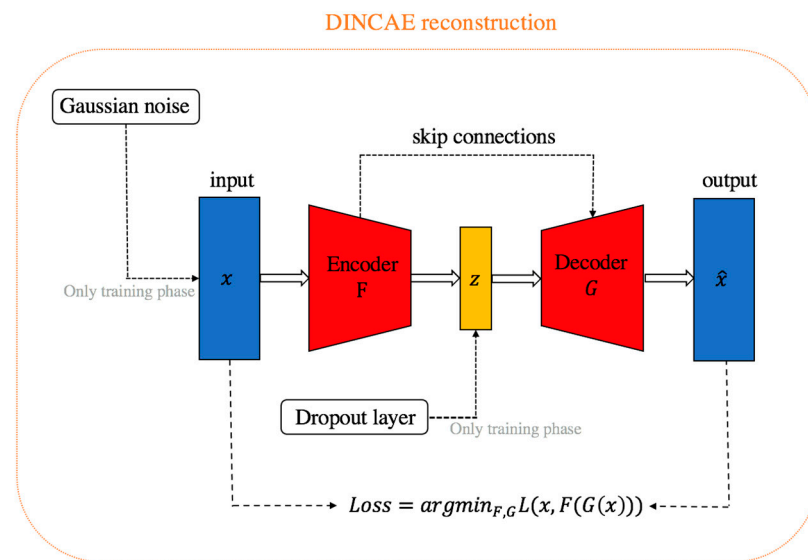

Figure 4. Architecture of the DINCAE model.

The input layer was used to receive the results of the data preprocessing. The encoder module extracted the key features and retained important information. Convolution layers play a key role in DINCAE: they can compress information in the encoder module and consist of the convolutional kernel and pooling layer. Fully connected layers are necessary modules in a feed–forward convolution neural network. They aim to combine all extracted features nonlinearly, which enhances the model fitting ability. In this work, the encoder module contained two full connection layers, which first reduced the input data dimensions from N to N/5 and then brought them back to N. The decoder module, as the encoder counterpart, was composed of convolution layers and an interpolation layer and up–sampled the results of the encoder module. The skip–connection structure, which was inspired by UNet [30], is vital for better prompting reconstruction. It can capture small–scale features that are lost in the encoding layers and fully connected layers and provide more useful information for interpolation. The output size was $480 \times 384 \times 2$, including the estimated Chl–a value and the expected error variance. The encoder module and decoder module both had five layers when their filter sizes were both standardized to 3×3. The rebuilt Chl–a value Chl–a$_{ij}$ and the corresponding error variance $\hat{\sigma}_{ij}$ were defined as:

$$\hat{\sigma}_{ij}^2 = \frac{1}{\max\left(\exp\left(\min\left(T_{ij1}, \gamma\right)\right), \delta\right)} \tag{6}$$

$$Chl - a_{ij} = T_{ij2}\hat{\sigma}_{ij}^2 \tag{7}$$

where $\gamma = 10$ and $\delta = 10^{-3}(\text{mg} \cdot \text{L}^{-1})^{-2}$, and T_{ij1} and T_{ij2} are the estimated value and error variance for the DINCAE output layer, respectively. When a neural network is trained by iteration, it is common to observe overfitting problems, which means that the model can fit the trained dataset well but works poorly with the test dataset. Some strategies can be adopted to avoid the overfitting phenomenon effectively:

1. During the training phase, Gaussian–distributed noise can be added to the input data. With the random noise, the model can address the overfitting problem and its robustness can be boosted;
2. Dropout [31], which makes the neural units randomly inactive in a fully connected layer, has proven its effectiveness in solving the overfitting issue in many studies [32]. Similar to the Gaussian–distributed noise, the above method was only implemented in the training phase and disabled during the reconstruction phase;
3. The activation function, which is used to enhance the nonlinear fitting capacity of a model, is widely applied after the fully connection layer and convolution layer. The formula for the rectified linear unit (Relu) is as follows:

$$f(x) = \begin{cases} x, (x \geq 0) \\ 0, (x < 0) \end{cases} \tag{8}$$

Based on the Relu function shown in Equation (8), the Leaky Relu [33], which is applied in DINCAE, can prevent gradients from falling too fast. The difference between the Relu and Leaky Relu functions is that, when the input data values are less than zero, the output of the Leaky Relu is αx (α is a hyperparameter and was set to 0.2 in this study), which still retains some gradient information. The code is available at https://github.com/gher--uliege/DINCAE accessed on 30 March 2022, and the work environment was Python 3.6 and Ubuntu 18.04 with GeForce RTX 3090.

2.5.3. BME

BME is an approach based on spatiotemporal random field theory [34] and information entropy that does not make any assumptions regarding the homogeneity of the spatial distribution, the normality of the underlying probability laws or the linearity of the estimator. It can integrate multiple data formats, including hard data and soft data, to improve its prediction accuracy. In accordance with the overall flowchart shown in Figure 5, the BME framework was divided into three stages: the prior stage, meta prior stage and posterior stage.

(a) Prior stage. The purpose of the prior stage was to obtain the prior probability density function by using various types of knowledge, such as physical laws, scientific theory, spatiotemporal covariance models, etc. Based on the general knowledge, the constraint equation can be described as the following:

$$\overline{g_a}(x) = \int g_a(x) f_G(x) dx \tag{9}$$

In the equation, a is the number of constraint conditions, g_a is the general knowledge function and f_G is the prior probability density function. In this study, the general knowledge was the mean value and the covariance function $c_X(\boldsymbol{p}, \boldsymbol{p}')$, shown in the following equation:

$$c_X(\boldsymbol{p}, \boldsymbol{p}') = E[X(\boldsymbol{p}) - E(X(\boldsymbol{p}))][X(\boldsymbol{p}') - E(X(\boldsymbol{p}'))] \tag{10}$$

where $\boldsymbol{p}$ denotes a spatiotemporal point and $X(\boldsymbol{p})$ is a collection of individual random variables.
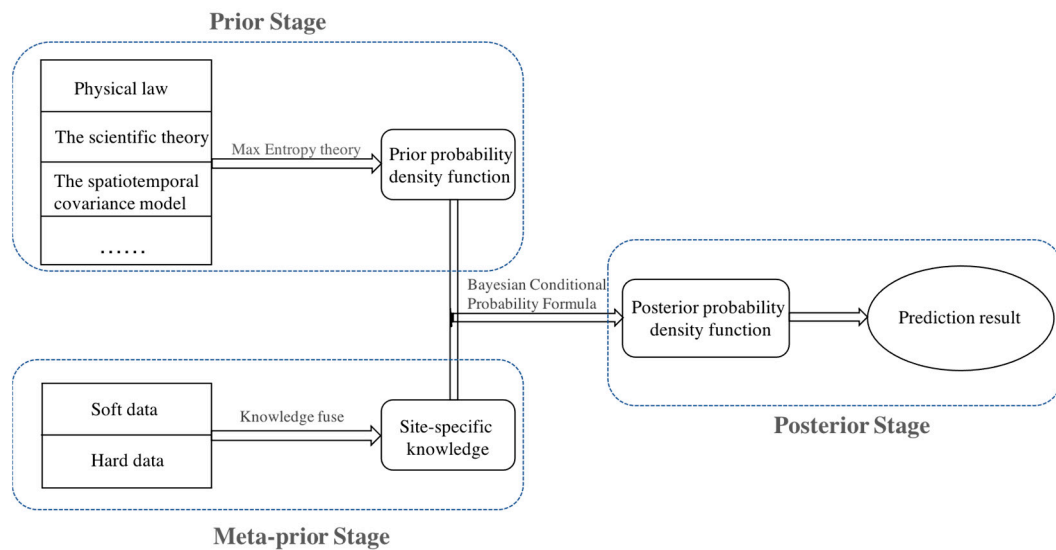
Figure 5. Schematic flowchart for BME.

(b)   Meta prior stage. The work in this stage comprised data processing, including hard and soft data preparation. The soft data had inherent uncertainty, such as that in low–precision remote sensing data, and they used three kinds of data formats: the interval format, probability format and function format. In this study, Chl–a data from the Aqua satellite were considered as hard data and used for DINEOF and DINCAE modeling, while the soft data were generated from the DINCAE output data, the original Terra data and the original S–NPP data, as shown in Figure 6.
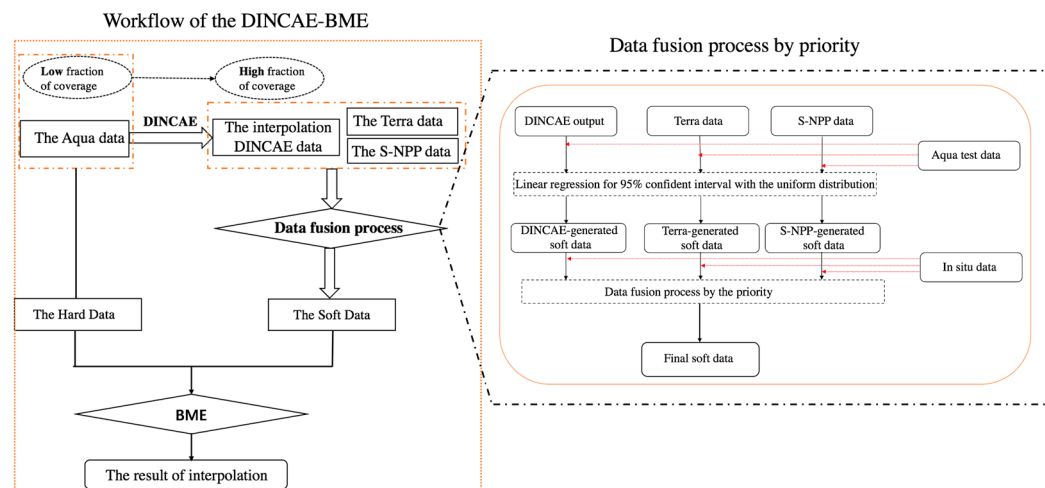

Figure 6. The DINCAE–BME framework. The BME absorbed the hard data and DINCAE–generated soft data.

The detailed process was as follows: the matched–up and paired data from the Aqua test dataset and the DINCAE output data/Terra data/S–NPP data were respectively used to build a linear model for calibration; then, all the DINCAE output data/Terra data/S–NPP data were introduced into the corresponding linear model to generate soft Chl–a data with a uniform–distribution probability density function by setting the upper and lower limits of the 95% confidence intervals of the linear models' outputs as the upper and lower limits of the uniform distribution. The linear regression results are shown in Table 2. Further, the in situ Chl–a data were matched and compared with the linear model output (or estimation) to determine the priority of the soft

data strategy chosen from among DINCAE–generated soft data, TERRA–generated soft data and S–NPP–generated soft data. Given that the R^2 values of the linear regression models for the in situ data and three kinds of estimations were 0.4661, 0.7246 and 0.6239, the priority for the choice of strategies for the soft data was set as Terra > S–NPP > output of DINCAE. Thus, the hard and soft data were prepared. In this study, the hard and soft data were both used for BME modeling.

Table 2. Linear regression results for Aqua test data with three data sources.

	N	Slope	Intercept (mg m^{-3})	R^2
DINCAE	1,304,562	0.8139	0.1227	0.6910
Terra	702,350	0.9290	0.2827	0.5500
S–NPP	817,499	1.4663	0.1227	0.6552

(c) Posterior stage. The main purpose of the posterior stage was to obtain the posterior probability distribution by making use of general knowledge, hard data and soft data. The process of determining the posterior probability distribution essentially involved solving the condition probability formula shown in the following equation:

$$f_K(\chi_k) = f_G(\chi_k | \chi_{data}) = \frac{f_G(\chi_k, \chi_{data})}{f_G(\chi_{data})} \tag{11}$$

Specifically, the posterior probability distribution formula can be described as follows when the soft data are in the probability format:

$$f_K(\chi_k) = \frac{\int_I f_G\left(\chi_{hard}, \chi_{soft}, \chi_k\right) dF\left(\chi_{soft}\right)}{\int_I f_G\left(\chi_{hard}, \chi_{soft}\right) dF\left(\chi_{soft}\right)} \tag{12}$$

In conclusion, it is obvious that the posterior stage used soft and hard data to solve the unknown coefficients in $f_G(\chi_{map})$, which is similar to the training stage in the machine learning model. STAR–BME was used in this work, which can be downloaded from https://stemlab.bse.ntu.edu.tw/. The program run for STAR–BME relied on QGIS software, which can be downloaded from https://www.qgis.org/en/site/. The work environment for STAR–BME was Windows 10.

2.5.4. DINCAE–BME Framework

This study proposed a DINCAE–BME framework to improve accuracy and coverage for Chl–a by using remote sensing data. The core idea was to integrate DINCAE–generated soft data with a BME methodology, and the method can borrow the capacity for coverage improvement from DINCAE and for accuracy modeling with uncertain data sources from BME. The workflow for the use of DINCAE–BME for Chl–a estimation is shown in Figure 6, and the details can be found in this section.
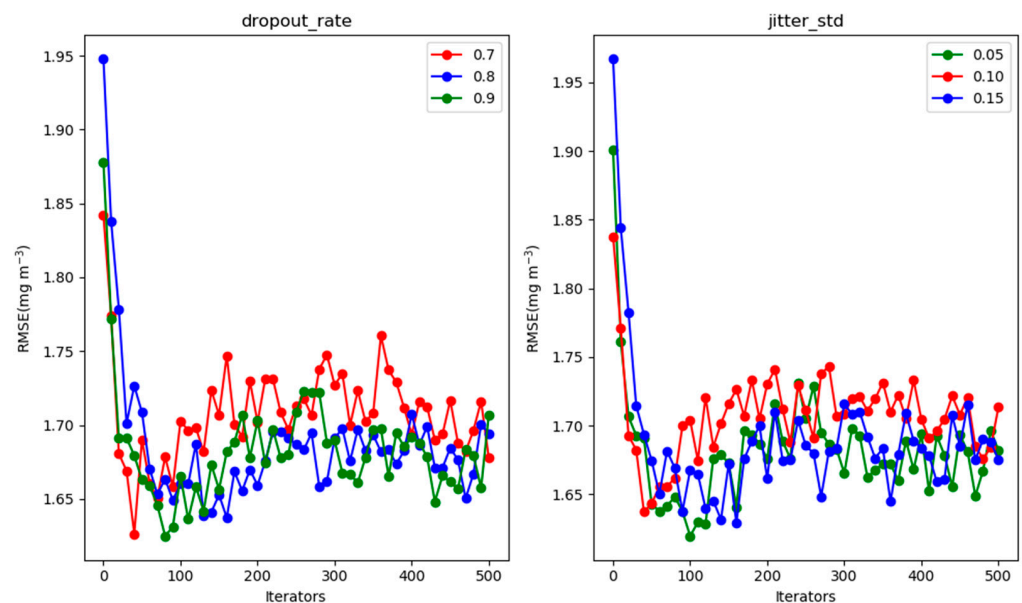
3. Results

3.1. Hyperparameter Experiment

In order to search for the optimal hyperparameter for the model, around 1% of the training data were split off to test the performance of the model with different hyperparameters. The number of EOFs was set within the range of 1 to 6 for DINEOF reconstruction, minimizing the expected error to 3.6089 mg m^{-3} when the EOF mode was 3. The results with different EOF modes for the DINEOF reconstruction process are shown in Table 3. Convergence was achieved for EOF–1, EOF–2, EOF–3, EOF–4, EOF–5 and EOF–6 after 66, 83, 78, 75, 99 and 300 iterations, respectively. In addition, the total variance in the reconstructed matrix was 7.4069 mg m^{-3}, while the total variance in the initial matrix was 17.6303 mg m^{-3}.

Table 3. The expected error with DINEOF for different EOF modes.

EOF Mode	Expected Error (mg m^{-3})	Iterations	Convergence Achieved (10^{-3})
1	3.7971	66	0.9864
2	3.6473	83	0.9815
3	3.6089	78	0.9867
4	3.6710	75	0.9983
5	3.7443	99	0.9959
6	3.8118	300	1.0510

DINCAE is a method based on deep learning that is associated with several hyperparameters. Consequently, the optimal combination of the hyperparameters in this work had to be determined using a step–by–step methodology. The jitter std parameter was chosen from (0.05, 0.10, 0.15) and the dropout rate was chosen from (0.7, 0.8, 0.9). During the training, the expected error between the reconstruction data and the roughly 1% split of test data was computed every ten epochs to evaluate the performance of the model. Figure 7 shows that the expected error curve exhibited a sharp decrease at the beginning and then began to converge and remained stable with slight fluctuations. The optimal combination was jitter std = 0.05 and dropout rate = 0.9 in this work.

Figure 7. Hyperparameter–finding experiments for DINCAE. The optimal dropout rate was chosen from (0.7, 0.8, 0.9) and the optimal jitter std was chosen from (0.05, 0.10, 0.15).

3.2. Cross–Validation Result

The comparative hold–out cross–validation results for the different methods are shown in Table 4 and the "validation data" are described in Section 2.4. For the DINCAE–BME method, the MAE and RMSE reached 0.4682 mg m^{-3} and 1.8824 mg m^{-3}, respectively. For the pure DINCAE, the MAE and RMSE were 0.7147 mg m^{-3} and 2.9860 mg m^{-3}. It was obvious that the hold–out cross–validation result for DINCAE–BME was better than that for the other methods. Furthermore, it was evident that the performance of DINCAE was better than that of DINEOF, which proved that DINCAE is more suitable for reconstructing Chl–a data.

Table 4. Results of hold–out cross–validation with different methods.

Method	MAE (mg m^{-3})	RMSE (mg m^{-3})
DINEOF	2.9710	4.9573
DINCAE	0.7147	2.9860
DINCAE−BME	0.4682	1.8824

3.3. Validation with the In Situ Data

It was crucial to validate the methods' performance with more convincing data in order to further illustrate the methods' effectiveness. A total of 242 in situ samples were matched to perform validation with the reconstruction results. The results' coverage percentages with DINEOF, DINCAE and DINCAE–BME were 78.4%, 99.7% and 100%, respectively. Days with no remote sensing data were not interpolated and were not considered in this study. In total, 156 points were matched with DINCAE–BME, while the match–point counts with DINCAE and DINEOF were 152 and 74. Before undertaking the evaluation comparing the in situ and construction results, the 156 in situ samples were split into months to analyze their distribution. Figure 8 shows the distribution of sample counts for each month, and there were more points in spring and winter than in autumn.
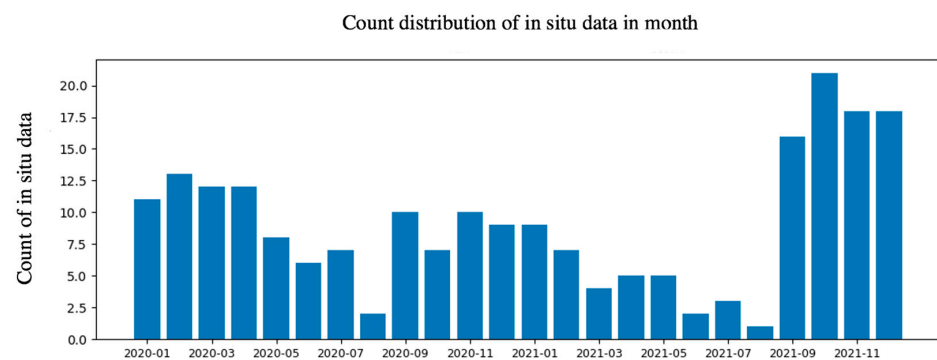
Figure 8. Data count distribution for the 156 in situ data samples.

A Comparison between the reconstructed data from the different methods and the matched in situ Chl–a measurements IS shown in Figure 9. The fitted line for DINCAE–BME shows that it had the best performance, with R^2, slope and intercept values of 0.6225, 1.1667 and −0.0276 mg m^{-3}, respectively. Compared to the other methods, DINCAE–BME also had lower RMSE and MAE values of 0.6196 mg m^{-3} and 0.3461 mg m^{-3}, as shown in Table 5, indicating that the reconstructed data from DINCAE–BME were closer to the in situ Chl–a measurements.
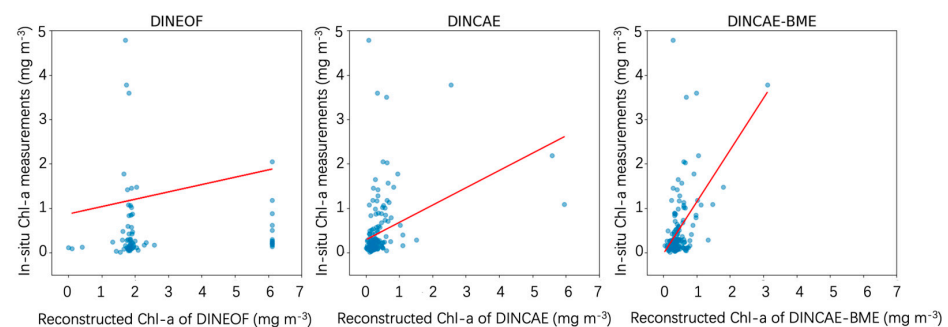

Figure 9. Comparison of the reconstructed data from different methods and in situ data. The match counts for DINEOF, DINCAE and DINCAE–BME were 74, 152 and 156.

Table 5. Relationship between reconstructed values and corresponding in situ data.

Number of Matched In Situ Data Samples	Method	RMSE (mg m^{-3})	MAE (mg m^{-3})	R^2	Slope	Intercept (mg m^{-3})
74	DINEOF	1.9580	1.2645	0.2871	0.1655	0.8682
152	DINCAE	0.8117	0.3797	0.4661	0.3900	0.1448
156	DINCAE–BME	0.6196	0.3461	0.6225	1.1667	−0.0276

3.4. Reconstruction Statistics

Figure 10 shows the mean values for the reconstruction of Chl–a data from the DINCAE–BME framework, and the coverage of the reconstruction results in the study area reached 100%. In order to describe its spatial distribution well, the results of the reconstruction are split into four typical regions in Figure 10: (a) the Gulf of Oman; (b) the Gulf of Khambhat; (c) the deep–sea region; and (d) the coastal region. Firstly, it was noticed that the (a) region, indicated by the square box in Figure 10, showed high Chl–a values and that some of the points' values were kept above 2.0 mg m^{-3}. The situation in the (a) region has also been observed in other studies [7,35]. Secondly, the (b) region showed higher activity than the (a) region; in contrast, Chl–a values in the (c) region were relatively low, with the values of most of the points remaining below 1.0 mg m^{-3}. Finally, it was observed that the (d) region was the region of high productivity; the Chl–a values of some of the points were kept above 4.0 mg m^{-3} and the overall trend was that the values gradually decreased from near the shore to the sea. This was mainly attributed to terrigenous input. Frequent human activities and the continuous discharge of sewage into the ocean due to the needs of industrial development have led to the eutrophication of water bodies, resulting in the proliferation of algae and the phenomenon of high Chl–a values [36].
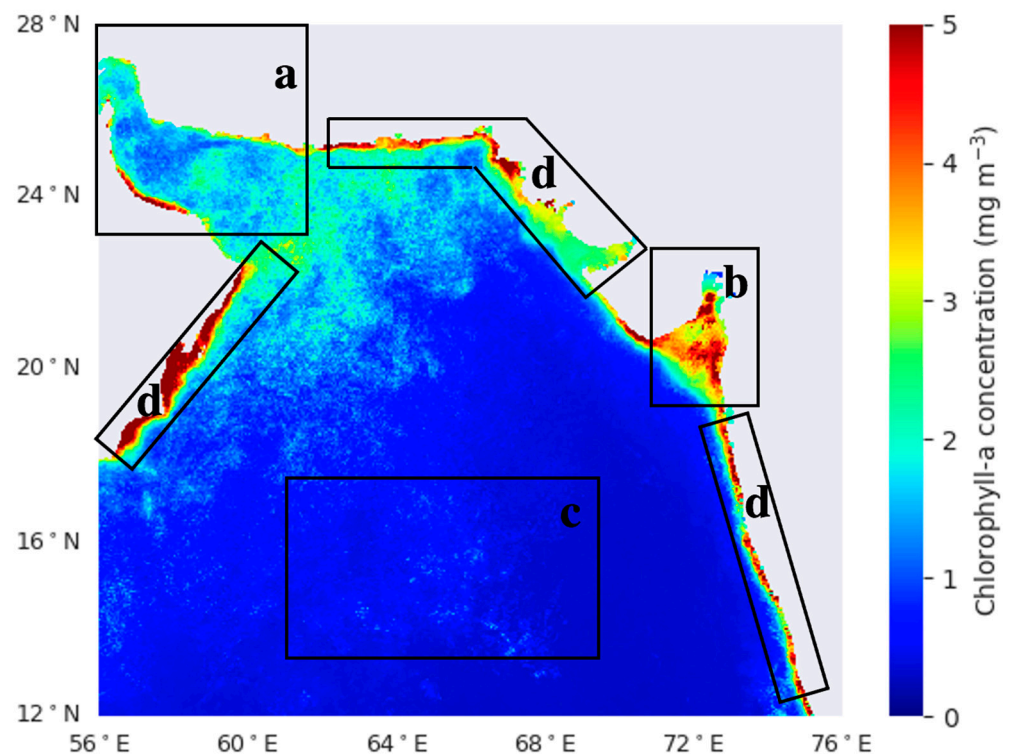

Figure 10. Mean values for the DINCAE–BME reconstruction results in a time series. (a) Region representing the Gulf of Oman; (b) region representing the Gulf of Khambhat, which is near India; (c) region representing the deep–sea region of the NAS; (d) region representing the coastal region.

In order to further analyze the spatiotemporal distribution of Chl–a and the influence factors, the reconstruction results were split into months to generate monthly average value results from 2020 and 2021, as shown in Figures 11 and 12, respectively. Specifically, the Chl–a values in the (a) region showed a clear increasing trend from January to March; in February, especially, the values showed an explosive increase and peaked. After March, they began to gradually decline. This temporal trend was similar to that found in other studies [37,38]. The situation in the (b) region was similar to that in the (a) region; Chl–a values in the (c) region also maintained a trend of increasing gradually from January to March and then began to decrease in their temporal distribution. As the overall mean was low, the temporal distribution in this region was not very obvious compared to the results from the Gulf regions. Finally, the Chl–a values in the (d) regions remained very high and did not show a clear temporal trend.

The temporal and spatial distribution characteristics of Chl–a in 2021, shown in Figure 12, were comparable to those in 2020. It is worth noting that the peak Chl–a values in the (a) region advanced to February, and the time trend changed to an upward trend from December to February that began to decline gradually after February. Moreover, compared with 2020, the peak area was smaller and the concentration was higher, with the majority of points dropping significantly in March and basically remaining below 1.0 mg m^{-3} for the rest of time. The Chl–a values in the (b) and (c) regions were consistent with the previous time trends. It was found that the Chl–a values in the (d) regions remained in a relatively stable temporal distribution with high values.
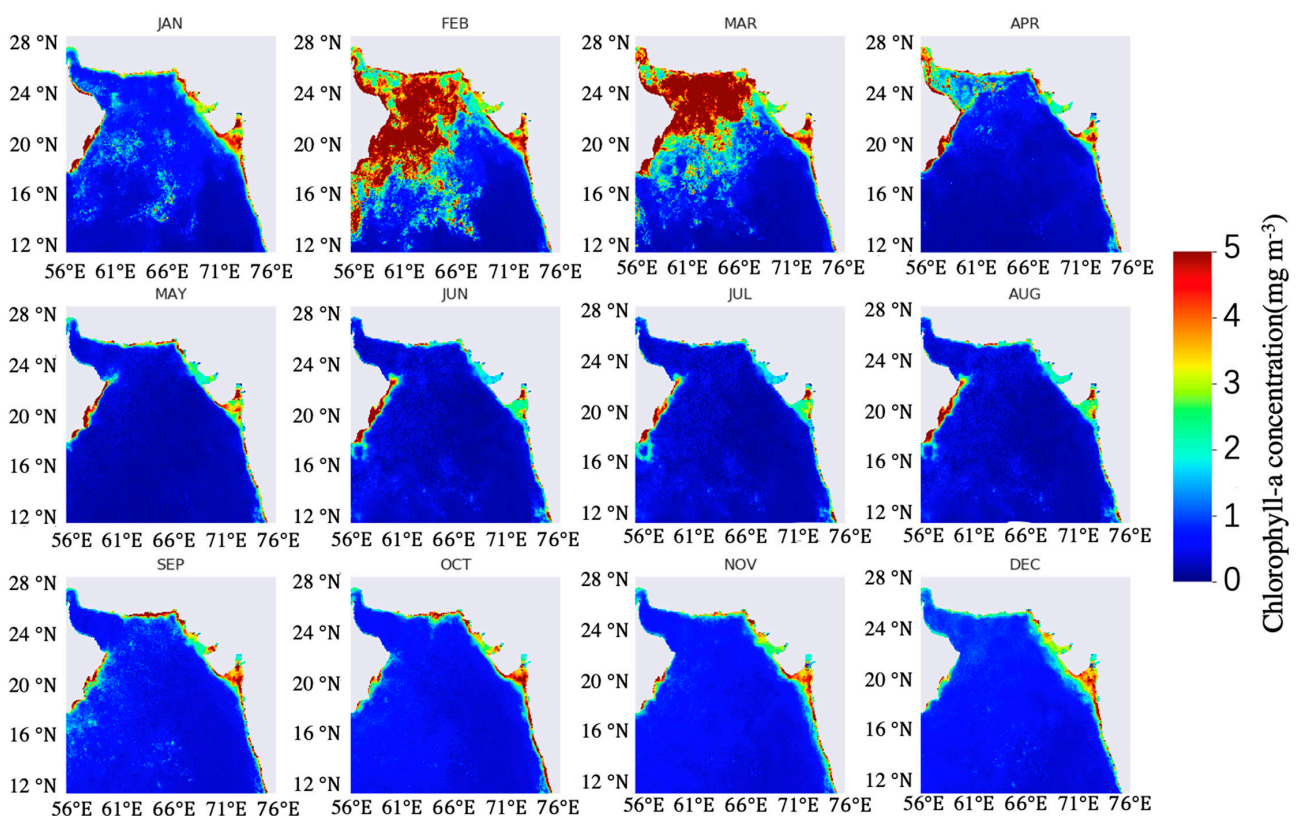

Figure 11. Monthly interpolation results for Chl–a with the DINCAE–BME from January to December 2020.
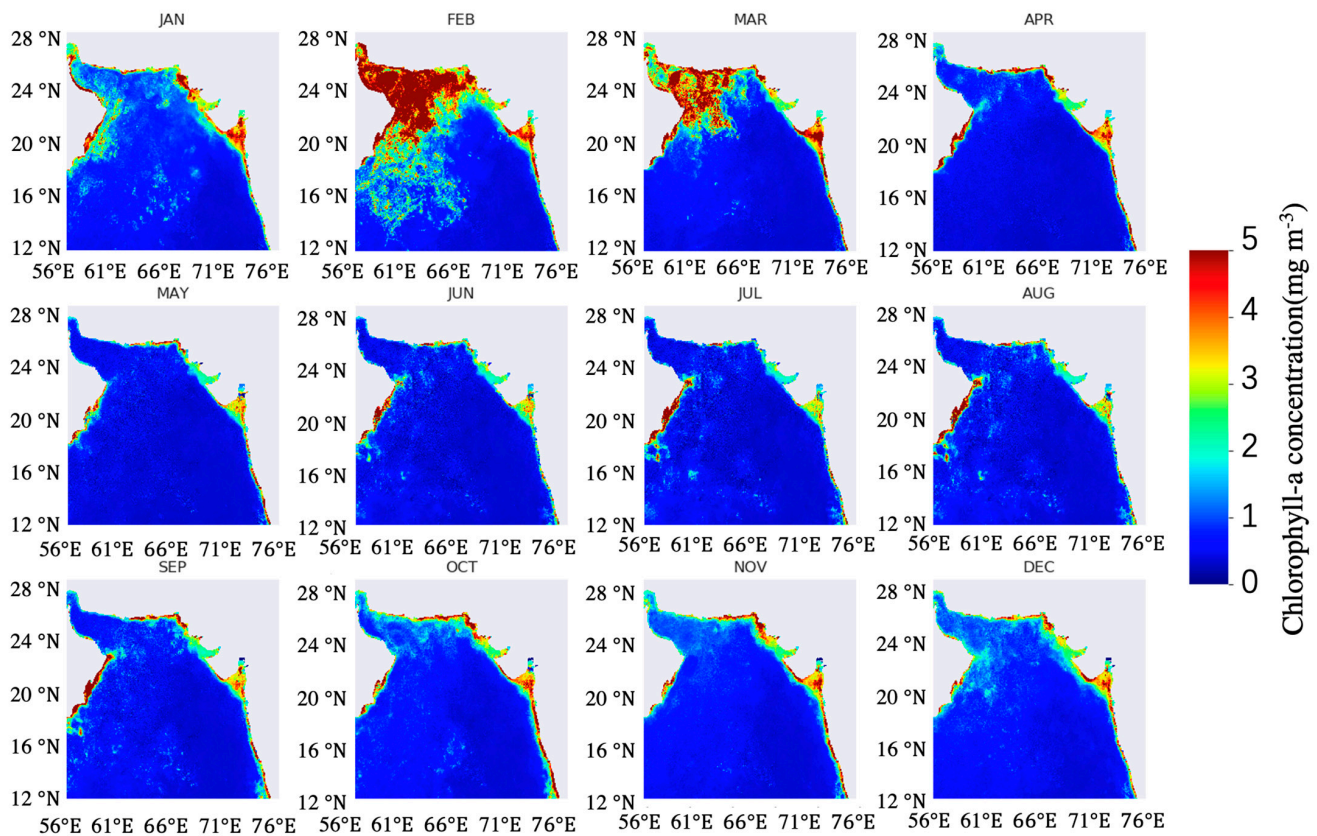
Figure 12. Monthly interpolation results for Chl–a with the DINCAE–BME from January to December 2021.

4. Discussion

4.1. Methodology

In this study, the DINCAE and BME methods were combined as a synthetic spatiotemporal interpolation framework to reconstruct the missing Chl–a remote sensing data for the NAS. The DINCAE method improved the average Chl–a data coverage rate from 17% to almost 100%. However, the model process for DINCAE only focused on models' own single data sources without considering other source information, such as Terra data and S–NPP data, which limited the accuracy of the interpolation results. Compared to traditional interpolation methods, such as the Kriging method, BME has a certain theoretical support and the ability to absorb uncertain soft data from other sources. As a nonlinear interpolator, BME optimizes the use and retention of various data information sources to further improve interpolation accuracy on the base of the DINCAE method.

The main benefit of this study framework is the strong generalizability and inclusiveness resulting from the fact that the DINCAE method can provide high–quality soft data and the BME method can effectively fuse the hard data and soft data, making the two methods complement each other in producing a more accurate result. DINCAE 2.0 [15], on the basis of DINCAE, reinforces the skip–connection structure, which was inspired by UNet [30], and employs deep supervision loss to better guide the encoder module learning. Although DINCAE 2.0 performed better in the multivariate reconstruction task than DINCAE, it still lacks the capacity to fuse multi–source data and information, which is the core advantage of the DINCAE–BME framework. In addition, the reinforced skip–connection structure with the concatenate style has higher requirements for memory and computing power, making it difficult to implement in the case of large–scale data. DINEOF–BME [23] and GLM–BME [39] are other attempts to extend the BME framework, achieving state–of–the–art results in some specific tasks. However, the first part in these methods only considers the simple linearity of spatiotemporal points, which further limits the perfor-

mance of BME. The DINCAE–BME framework utilizes the strong power of deep learning to capture more nonlinear relationships and retain small–scale features of data to provide more confident soft data.

In conclusion, the DINCAE–BME framework achieves excellent results in the interpolation of missing values because of the effective combination of deep learning models and geostatistical methods. The first part of the framework, DINCAE, is used to improve the coverage of the soft data, taking into account the accuracy of interpolation and data coverage. The second part of the framework, BME, is used for the fusion of multi–source information, learning the knowledge from hard and soft data and further improving the accuracy of the interpolation. The effectiveness of the framework was effectively demonstrated in cross–validation experiments and comparison experiments with in situ data. The DINCAE–BME framework proposed in this work is a general framework for interpolation and can also be applied in various other areas where there is the same problem of low coverage of data. Compared with other methods, the core advantage of DINCAE–BME is that it does not require further auxiliary variable information, which is crucial in some fields. For example, for interpolation research tasks relating to SST, seawater salinity and aerosol concentration, there are situations where other auxiliary variables are unavailable and the DINCAE–BME framework can be considered as one potential solution.

4.2. Product Analysis

By utilizing the precise interpolation capacity of the DINCAE–BME method, the reconstruction results for Chl–a in 2020 and 2021 were obtained in the study area, increasing the data coverage percentage to 100%. From the perspective of spatial distribution, it was observed that Chl–a values indicated high productivity in the (d) regions, where some points were maintained around 4.0 mg m^{-3} and the overall trend was a gradual decrease from near the shore to the sea, which may have been caused by terrigenous input. Due to frequent human activities and the emission and diffusion of pollutants, the eutrophication of coastal seawater is leading to large outbreaks of phytoplankton, resulting in high Chl–a values. The (a) and (b) regions also remain active situations, with Chl–a values at most points above 2.0 mg m^{-3}, while the Chl–a values in the (c) region were relatively low. The high concentration of nutrients in the (c) region combined with the influence of various factors resulted in the explosive increase in Chl–a values in the (a) region from February to March [40]. Chl–a values in the (b) region were similar to those in the (a) region in terms of temporal distribution. The difference was that Chl–a values were higher in the (b) region than in the (a) region most of the time. This was mainly because the (b) region was located near Mumbai, and human activities are more frequent than in the (a) region. More importantly, the (a) region was situated with the Arabian Sea to the east and the Persian Gulf to the west, which is a relatively connected gulf. The (b) region was more closed than the (a) region, and there was relatively less material exchange with the deep–sea area, which may have had an influence on Chl–a.

From the perspective of the temporal distribution, the reconstruction results showed significant seasonal changes, which may have been influenced by wind speed, tide [35] and SST [41]. The monsoon activity from January to March was more frequent and the wind speed was high, which led to sufficient exchange between the seawater surface and bottom layer and, at the same time, brought nutrients to the surface layer, promoting the growth and reproduction of phytoplankton [42] and thus affecting the Chl–a concentration. In addition to the influence of the monsoon, the thickness of the mixed layer could have allowed the nutrient–rich lower water to surge into the upper layer, and during this period, the injection of nutrients resulted in an increase in Chl–a values in the NAS [43]. In addition, seasonal changes in SST can cause water stratification or mixing, which, in turn, may have had a certain impact on the seasonal distribution of Chl–a [36].

This may be attributed to the decrease in the upper water temperature in winter leading to the densification effect in the upper water body, which causes instability in this body. In order to further explore the variation in Chl–a in the Arabian Sea area from 2020

to 2021, the difference between the reconstruction results for 2020 and 2021 is shown in Figure 13. The results showed that Chl–a values maintained a stable range from April to December in the (c) region, with most of the points approaching zero. However, the situation was different that the (a) region, which remained slightly yellow, indicating that the Chl–a values underwent a small increase from 2020 to 2021 in the autumn and winter. It can be observed from the figures that the growth situation increased over time, with a change from yellow to red in some areas. It was previously mentioned that Chl–a values demonstrated an explosive increase and peaked in February and March of both 2020 and 2021, as shown in Figures 11 and 12. The low SST and high light intensity were also conducive to the growth and reproduction of marine organisms, which further contributed to the increased in Chl–a concentration [44]. The cool and dry northeast monsoon in the NAS enhances evaporation from the ocean surface water. The decrease in SST is mainly due to evaporation, which may cause an increase in salinity, further leading to convective overturning and initiating deepening of the mixed–layer depth [44]. The moderate wind could have enhanced the process of mixing of nutrients toward the surface [45]. It is worth focusing on February and March in Figure 13. In February, the area of the (a) region is red, which means increasing Chl–a values, while the other areas of the sea are blue, which means decreasing Chl–a values. What is interesting is that the explosive increase in Chl–a values disappeared in March, being replaced by a situation in which Chl–a values were decreasing across almost the entire Arabian Sea. Furthermore, the Chl–a in the coastal region showed an obvious decreasing trend. It is worth mentioning that, from 2020 to 2021, COVID–19 gradually began to spread around the world. The spread of COVID–19 led to a significant reduction in human activities, affecting terrigenous input in the coastal region [46] and further reducing Chl–a values in coastal areas. Figure 14 shows the temporal distribution for 2020 and 2021, ignoring the influence of spatial distribution. First of all, looking at the comparison results for 2020 and 2021, the Chl–a values in 2021 were generally lower than in 2020, especially in the outbreak period of February and March, with the gap in March being very obvious. After May, the overall difference was not significant. From the perspective of seasonal changes, the overall Chl–a values in the Arabian Sea were higher in winter than in summer and higher in autumn than in spring.
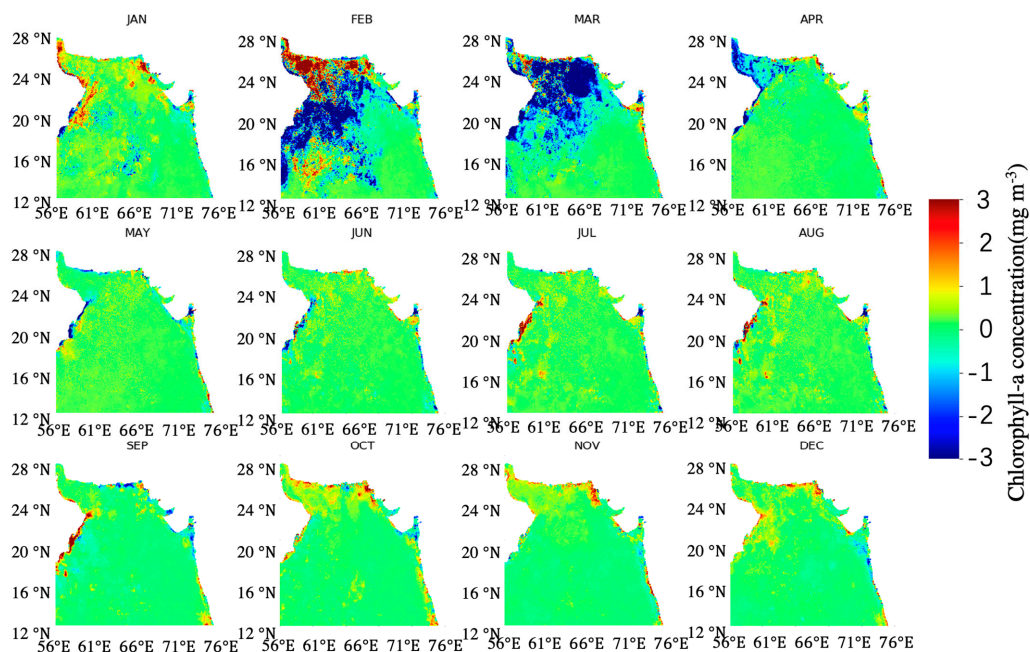

Figure 13. Difference in monthly interpolation results for Chl–a obtained with DINCAE–BME from January to December, generated by subtracting 2020 results from 2021 results in months.
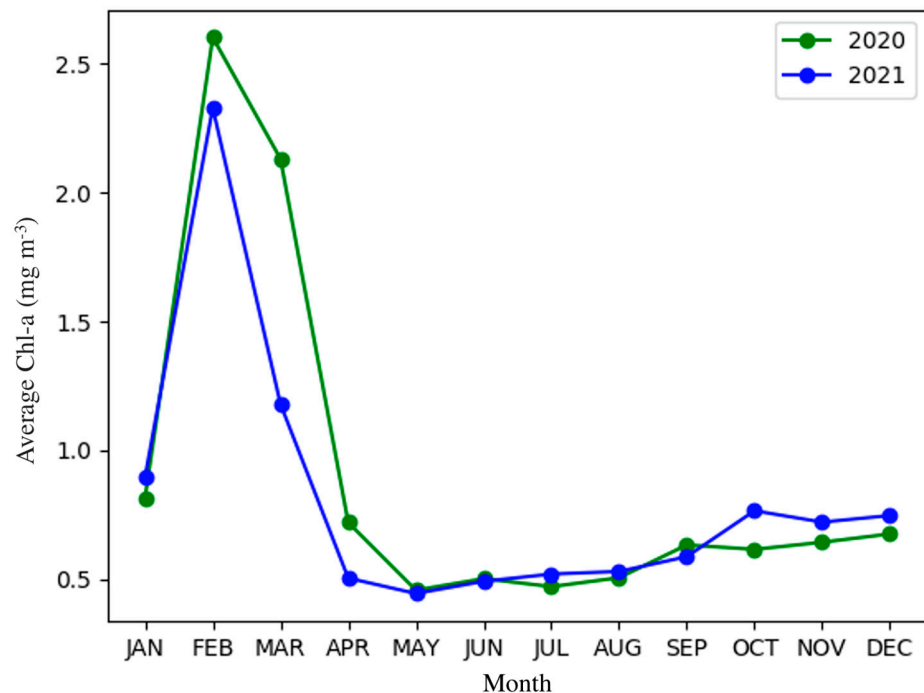
Figure 14. Monthly average Chl–a values in 2020 and 2021.

5. Conclusions

In this paper, the DINCAE method and BME method were combined to interpolate Chl–a data from 2020 to 2021 in the geographical range 12–24° N and 56–76° E. The proposed DINCAE–BME method was evaluated using both validation data and in situ data. Compared with in situ Chl–a measurements, the MAE, RMSE and R^2 values for the interpolated products were 0.3461 mg m^{-3}, 0.6196 mg m^{-3} and 0.6225 mg m^{-3}, respectively. The results showed that DINCAE–BME had smaller errors compared to pure DINCAE, which supported the utility of combining the DINCAE and BME procedures. Furthermore, the DINCAE method and the DINEOF method were evaluated using validation data and in situ data. The MAE and RMSE computed using in situ data and the reconstruction results from DINCAE were 0.3797 mg m^{-3} and 0.8117 mg m^{-3}, while the two indicator values for DINEOF were 1.2645 mg m^{-3} and 1.9580 mg m^{-3}. It was found that DINCAE performed better than DINEOF in the task of Chl–a interpolation in the NAS because of its capacity to capture nonlinear and mutated relationships. In addition, the results showed that the DINCAE–BME framework increased the coverage of Chl–a data from 17% to 100% with a high–accuracy interpolation value. The spatial distribution of the reconstruction results showed that the Chl–a values in the coastal region were high because of human activity, and those in the gulf region were higher than those in the deep–sea region. The temporal distribution of the results showed that values for Chl–a in winter were higher than in summer, and the values in autumn were higher than in spring. The difference between the results for 2020 and 2021 verified the conclusion that COVID–19 had an impact on the marine environment by limiting human activity. Furthermore, the inclusiveness of BME was further demonstrated by its capacity to absorb the multi–source information.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; validation, X.Y.; formal analysis, X.Y.; investigation, X.Y.; resources, X.Y.; data curation, X.Y.; writing—original draft preparation, X.Y; writing—review and editing, X.Y., Z.G., Y.J., J.H., J.W. and J.Y. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   He, J.; Chen, Y.; Wu, J.; Stow, D.A.; Christakos, G. Space–Time Chlorophyll–a Retrieval in Optically Complex Waters That Accounts for Remote Sensing and Modeling Uncertainties and Improves Remote Estimation Accuracy. *Water Res.* **2020**, *171*, 115403. [CrossRef]
2.   Kasprzak, P.; Padisák, J.; Koschel, R.; Krienitz, L.; Gervais, F. Chlorophyll a Concentration across a Trophic Gradient of Lakes: An Estimator of Phytoplankton Biomass? *Limnologica* **2008**, *38*, 327–338. [CrossRef]
3.   Johan, F.; Jafri, M.Z.; Lim, H.S.; Wan Maznah, W.O. Laboratory Measurement: Chlorophyll–a Concentration Measurement with Acetone Method Using Spectrophotometer. In *2014 IEEE International Conference on Industrial Engineering and Engineering Management*; IEEE: New York, NY, USA, 2014.
4.   Hamilton, R.J.; Sewell, P.A. *Introduction to High Performance Liquid Chromatography*; Springer: Dordrecht, The Netherlands, 1982; pp. 1–12.
5.   Xing, X.G.; Zhao, D.Z.; Liu, Y.G.; Yang, J.; Shen, H. Progress in fluorescence remote sensing of chlorophyll–a. *J. Remote Sens.* **2007**, *11*, 137–144.
6.   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]
7.   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]
8.   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]
9.   Hilborn, A.; Costa, M. Applications of DINEOF to satellite–derived chlorophyll–a from a productive coastal region. *Remote Sens.* **2018**, *10*, 1449. [CrossRef]
10.   Jayaram, C.; Priyadarshi, N.; Pavan Kumar, J.; Udaya Bhaskar, T.V.S.; 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. [CrossRef]
11.   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]
12.   Ji, C.; Zhang, Y.; Cheng, Q.; Tsou, J.; Jiang, T.; Liang, X.S. Evaluating the Impact of Sea Surface Temperature (SST) on Spatial Distribution of Chlorophyll–a Concentration in the East China Sea. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *68*, 252–261. [CrossRef]
13.   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]
14.   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]
15.   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]
16.   Luo, X.; Song, J.; Guo, J.; Fu, Y.; Wang, L.; Cai, Y. Reconstruction of Chlorophyll–a Satellite Data in Bohai and Yellow Sea Based on DINCAE Method. *Int. J. Remote Sens.* **2022**, *43*, 3336–3358. [CrossRef]
17.   Barth, A.; Alvera–Azcarate, A.; Troupin, C.; Beckers, J.-M.; Van der Zande, D. Reconstruction of Missing Data in Satellite Images of the Southern North Sea Using a Convolutional Neural Network (Dincae). In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*; IEEE: New York, NY, USA, 2021.
18.   Houyoux, A. Reconstruction of Missing Data in HF Radar Observations Using the Convolutional Autoencoder DINCAE. Master's Thesis, University of Liège, Liège, Belgium, 2021.
19.   Kostopoulou, E. Applicability of Ordinary Kriging Modeling Techniques for Filling Satellite Data Gaps in Support of Coastal Management. *Model. Earth Syst. Environ.* **2020**, *7*, 1145–1158. [CrossRef]
20.   Hou, P.; Luo, Y.; Yang, K.; Shang, C.; Zhou, X. Changing Characteristics of Chlorophyll a in the Context of Internal and External Factors: A Case Study of Dianchi Lake in China. *Sustainability* **2019**, *11*, 7242. [CrossRef]
21.   He, J.; Christakos, G.; Wu, J.; Li, M.; Leng, J. Spatiotemporal BME Characterization and Mapping of Sea Surface Chlorophyll in Chesapeake Bay (USA) Using Auxiliary Sea Surface Temperature Data. *Sci. Total Environ.* **2021**, *794*, 148670. [CrossRef]
22.   Christakos, G. A Bayesian/Maximum–Entropy View to the Spatial Estimation Problem. *Math. Geol.* **1990**, *22*, 763–777. [CrossRef]
23.   Jiang, Y.; Gao, Z.; He, J.; Wu, J.; Christakos, G. Application and Analysis of XCO_2 Data from OCO Satellite Using a Synthetic DINEOF–BME Spatiotemporal Interpolation Framework. *Remote Sens.* **2022**, *14*, 4422. [CrossRef]

24. Gao, Z.; Jiang, Y.; He, J.; Wu, J.; Christakos, G. Bayesian Maximum Entropy Interpolation of Sea Surface Temperature Data: A Comparative Assessment. *Int. J. Remote Sens.* **2022**, *43*, 148–166. [CrossRef]

25. He, M.; He, J.; Christakos, G. Improved Space–Time Sea Surface Salinity Mapping in Western Pacific Ocean Using Contingogram Modeling. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 355–368. [CrossRef]

26. Lang, Y.; Christakos, G. Ocean Pollution Assessment by Integrating Physical Law and Site-specific Data. *Environmetrics* **2019**, *30*, e2547. [CrossRef]

27. Shafeeque, M.; George, G.; Akash, S.; Smitha, B.R.; Shah, P.; Balchand, A.N. Interannual Variability of Chlorophyll–a and Impact of Extreme Climatic Events in the South Eastern Arabian Sea. *Reg. Stud. Mar. Sci.* **2021**, *48*, 101986. [CrossRef]

28. Shi, W.; Wang, M. Phytoplankton Biomass Dynamics in the Arabian Sea from VIIRS Observations. *J. Mar. Syst.* **2022**, *227*, 103670. [CrossRef]

29. Lei, Q.; Luo, H.; Bai, L. Space–time dynamic changes of aerosols in the Arabian Sea and characteristics of chlorophyll a concentration in the sea area. *Chin. J. Ecol.* **2019**, *39*, 3110–3120.

30. Ronneberger, O.; Fischer, P.; Brox, T. U–Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

31. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

32. Cheng, G.; Peddinti, V.; Povey, D.; Manohar, V.; Khudanpur, S.; Yan, Y. An Exploration of Dropout with LSTMs. In *Interspeech 2017*; ISCA: Dublin, Ireland, 2017.

33. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.

34. Christakos, G. *Random Field Models in Earth Sciences*; Dover Publications: Mineola, NY, USA, 2012.

35. Gernez, P.; Doxaran, D.; Barillé, L. Shellfish Aquaculture from Space: Potential of Sentinel2 to Monitor Tide–Driven Changes in Turbidity, Chlorophyll Concentration and Oyster Physiological Response at the Scale of an Oyster Farm. *Front. Mar. Sci.* **2017**, *4*, 137. [CrossRef]

36. Zhang, H.; Qiu, Z.; Sun, D.; Wang, S.; He, Y. Seasonal and Interannual Variability of Satellite–Derived Chlorophyll–a (2000–2012) in the Bohai Sea, China. *Remote Sens.* **2017**, *9*, 582. [CrossRef]

37. Piontkovski, S.; Al–Azri, A.; Al–Hashmi, K. Seasonal and Interannual Variability of Chlorophyll–a in the Gulf of Oman Compared to the Open Arabian Sea Regions. *Int. J. Remote Sens.* **2011**, *32*, 7703–7715. [CrossRef]

38. Yoder, J.A. An Overview of Temporal and Spatial Patterns in Satellite–Derived Chlorophyll–a Imagery and Their Relation to Ocean Processes. *Elsevier Oceanogr. Ser.* **2000**, *63*, 225–238.

39. Mei, Y.; Li, J.; Xiang, D.; Zhang, J. When a Generalized Linear Model Meets Bayesian Maximum Entropy: A Novel Spatiotemporal Ground–Level Ozone Concentration Retrieval Method. *Remote Sens.* **2021**, *13*, 4324. [CrossRef]

40. Ghaemi, M.; Abtahi, B.; Gholamipour, S. Spatial Distribution of Nutrients and Chlorophyll a across the Persian Gulf and the Gulf of Oman. *Ocean Coast. Manag.* **2021**, *201*, 105476. [CrossRef]

41. Shalin, S.; Samuelsen, A.; Korosov, A.; Menon, N.; Backeberg, B.C.; Pettersson, L.H. Delineation of Marine Ecosystem Zones in the Northern Arabian Sea during Winter. *Biogeosciences* **2018**, *15*, 1395–1414. [CrossRef]

42. Yang, L. Research progress in determination of phytoplankton chlorophyll–a. *Sichuan Environ.* **2019**, *38*, 156–160.

43. Dey, S.; Singh, R.P. Comparison of Chlorophyll Distributions in the Northeastern Arabian Sea and Southern Bay of Bengal Using IRS–P4 Ocean Color Monitor Data. *Remote Sens. Environ.* **2003**, *85*, 424–428. [CrossRef]

44. Naqvi, A.; Narvekar, V.; Desa, E. Coastal biogeochemical processes in the north Indian Ocean (14, S–W). In *The Sea: Ideas and Observations on Progress in the Study of the Seas*; Robinson, A.R., Brink, K.H., Eds.; Wiley: Hoboken, NJ, USA, 2006; pp. 723–781.

45. Garcia, E.; Locarnini, A.; Boyer, P.; Antonov, I. Dissolved inorganic nutrients (phosphate, nitrate, silicate). *World Ocean. Atlas* **2013**, *4*, 25.

46. Braga, F.; Ciani, D.; Colella, S.; Organelli, E.; Pitarch, J.; Brando, V.E.; Bresciani, M.; Concha, J.A.; Giardino, C.; Scarpa, G.M.; et al. COVID–19 Lockdown Effects on a Coastal Marine Environment: Disentangling Perception versus Reality. *Sci. Total Environ.* **2022**, *817*, 153002. [CrossRef] [PubMed]