Neural Network Approaches to Reconstruct Phytoplankton Time-Series in the Global Ocean

Elodie Martinez, Anouar Brini, Thomas Gorgues, Lucas Drumetz, Joana Roussillon, Pierre Tandeo, Guillaume Maze, Ronan Fablet

▶ To cite this version:

Elodie Martinez, Anouar Brini, Thomas Gorgues, Lucas Drumetz, Joana Roussillon, et al.. Neural Network Approaches to Reconstruct Phytoplankton Time-Series in the Global Ocean. Remote Sensing, 2020, 12 (24), pp.4156. 10.3390/rs12244156 . hal-03126743

HAL Id: hal-03126743
https://hal.univ-brest.fr/hal-03126743v1

Submitted on 8 Oct 2021

remote sensing

MDPI

Letter

Neural Network Approaches to Reconstruct Phytoplankton Time-Series in the Global Ocean

Elodie Martinez [1,*], Anouar Brini [1], Thomas Gorgues [1], Lucas Drumetz [2], Joana Roussillon [1], Pierre Tandeo [2], Guillaume Maze [1] and Ronan Fablet [2]

[1] Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, University Brest-CNRS-IRD-Ifremer, 29200 Brest, France; anouar.brini@supcom.tn (A.B.); thomas.gorgues@ird.fr (T.G.); Joana.Roussillon@etudiant.univ-brest.fr (J.R.); Guillaume.Maze@ifremer.fr (G.M.)

[2] IMT Atlantique, Lab-STICC, UMR CNRS 6285, 29200 Brest, France; lucas.drumetz@imt-atlantique.fr (L.D.); pierre.tandeo@imt-atlantique.fr (P.T.); ronan.fablet@imt-atlantique.fr (R.F.)

* Correspondence: elodie.martinez@ird.fr

Received: 27 November 2020; Accepted: 16 December 2020; Published: 18 December 2020

Abstract: Phytoplankton plays a key role in the carbon cycle and supports the oceanic food web. While its seasonal and interannual cycles are rather well characterized owing to the modern satellite ocean color era, its longer time variability remains largely unknown due to the short time-period covered by observations on a global scale. With the aim of reconstructing this longer-term phytoplankton variability, a support vector regression (SVR) approach was recently considered to derive surface Chlorophyll-a concentration (Chl, a proxy of phytoplankton biomass) from physical oceanic model outputs and atmospheric reanalysis. However, those early efforts relied on one particular algorithm, putting aside the question of whether different algorithms may have specific behaviors. Here, we show that this approach can also be applied on satellite observations and can even be further improved by testing performances of different machine learning algorithms, the SVR and a neural network with dense layers (a multi-layer perceptron, MLP). The MLP outperforms the SVR to capture satellite Chl (correlation of 0.6 vs. 0.17 on a global scale, respectively) along with its seasonal and interannual variability, despite an underestimated amplitude. Among deep learning algorithms, neural network such as MLP models appear to be promising tools to investigate phytoplankton long-term time-series.

Keywords: phytoplankton time-series reconstruction; ocean color; neural networks; support vector regression; multi-layer perceptron; physical predictors

1. Introduction

Phytoplankton—the microalgae that populates the upper sunlit layers of the ocean—fuels the oceanic food web and regulates oceanic and atmospheric carbon dioxide levels through photosynthetic carbon fixation ([1,2]). Seasonal and inter-annual cycles of phytoplankton biomass are now relatively well characterized, thanks to the large amount of studies based on radiometric satellite observations (e.g., [3,4]). Since the launch of the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) in late 1997, satellite radiometric observations are continuous. However, 20 years of observations is still too short to thoroughly investigate decadal Chlorophyll-a concentration (Chl, a proxy of phytoplankton biomass) variations. The unavailability of global scale observations over a continuous time-series longer than two decades led the scientific community to rely on coupled physical–biogeochemical ocean modeling to investigate phytoplankton biomass decadal variability. While models are able to resolve seasonal to interannual biogeochemical variability to an ever-improving degree (e.g., [5,6]), they diverge in reproducing decadal observations, in particular phytoplankton regime shifts [7–9]. Consequently, it is still not possible on a global scale to clearly separate the phytoplankton long-term response to climate change from natural variability.

However, well-characterized decadal cycles of phytoplankton (in terms of biomass, community composition, and carbon fluxes) are crucial as (1) They can accentuate, weaken, or even mask the climate-related trends (the recent debate about the observed North Atlantic regional cooling in the context of climate change illustrates the crucial need for better understanding decadal variability [10]; (2) The observed changes in phytoplankton during decadal cycle warm phases may provide insights into how future climate warming-induced changes will alter carbon cycle and the marine food web.

The distribution of phytoplankton is strongly controlled by physical processes over a large part of the global ocean (e.g., [11–14]). Consequently, past Chl variations may be reconstructed from past physical environmental factors. To our knowledge only two studies have been performed to reconstruct surface Chl in order to investigate its decadal variability. The first one allowed the derivation of spatio-temporal surface Chl variations over the 1958–2008 period in the tropical Pacific [15]. This reconstruction used a linear canonical correlation analysis on sea surface temperature (SST) and sea surface height to improve the description of the Chl response to the diversity of observed El Niño events and decadal climate variations in the tropical Pacific. The second one investigated the ability of a non-linear statistical approach based on support vector regression (SVR) to reconstruct historical Chl variations on a global scale using selected surface oceanic and atmospheric physical variables from a numerical model as predictors [16]. The SVR method was able to reproduce trends as well as the main modes of the interannual Chl variability depicted by satellite observations in most regions. Changes observed by satellite Chl between the 1980s and 2000s were also qualitatively captured by the SVR. The main bias of this approach was to underestimate the amplitude of the Chl variations by a factor of two.

Here, we investigate how a multi-layer perceptron (MLP) may be more skillful than this SVR approach to reconstruct satellite Chl on a global scale. While, in [16], we used physical outputs from an ocean forced model to train the SVR and reconstruct surface Chl, here we choose to only use physical predictors from satellite observations and numerical atmospheric reanalysis. This choice is motivated by (i) our ultimate objective, which is to reconstruct Chl from physical observations (i.e., not relying on biogeochemical numerical models); and (ii) the use of the most realistic environmental conditions that those observations allow. However, those observations are mainly available through remotely sensed surface data (oceanic observations below the surface are indeed usually not accessible at large spatial-scales or interannual time-scales) and predictors are then limited to surface variables. With such limited 2D sampling, we are aiming at building a statistical model that may challenge more complex numerical models which simulate complex three-dimensional processes. Indeed, such models may not only strongly diverge in capturing Chl variations at a timescale of a decade but they are also not straightforward to run and require large computing resources.

2. Materials and Methods

2.1. Oceanic and Atmospheric Datasets

Phytoplankton needs light and nutrients to grow. Physical processes strongly control spatial distribution and time-variability of nutrient inputs in the upper-sunlit layer, and thus of phytoplankton over a large part of the global ocean. Thus, we make use of this physical (bottom-up) control to derive statistical models that relate several physical variables (predictors) to satellite Chl (output) as detailed in Table 1.

Table 1. Physical predictors, their relevance to Chl variability, the products used, and their resolution.

Proxy Used as Predictors	Relevance to Chl Variations and Associated References	Products	Spatio-Temporal Resolutions
SST	Vertical mixing and upwelling [17–20] Impacts on phytoplankton metabolic rates [21]	Reyn_SmithOIv2 SST dataset [22]	Monthly on a 1° × 1° spatial grid
Sea level anomaly	Thermocline/pycnocline depths [11,23,24]	Ssalto/Duacs merged product of CNES/SALP project [25]	Weekly on a 1/3° × 1/3° spatial grid
Zonal and meridional surface winds	Surface momentum flux forcing and vertical motions driven by Ekman pumping [20,26]	Atmospheric model reanalysis ERA interim 4 [27]	Every 5-days on a 0.25° × 0.25° spatial grid
Zonal and meridional surface total currents	Horizontal advective processes [4,28]	OSCAR unfiltered satellite product [29]	Every 5-days on a 0.25° × 0.25° spatial grid
Short-wave radiations	Photosynthetically active radiation	NCEP/NCAR Numerical reanalysis [30]	Daily on a 2° grid
Month (cos and sin)	Periodicity of the day of the year (day 1 is very similar to day 365 from a seasonal perspective) [31]		
Longitude (cos and sin) and latitude	Periodicity (longitude 0° = longitude 360°) [31]		

Chl was retrieved from the Ocean Colour–Climate Change Initiative (OC-CCI) from the European Space Agency (ESA, [32]). This product has generated global, ocean-colour products for climate research by merging observations from different sensors while attempting to reduce inter-sensor biases [33]. The V4.2 product was extracted on a 1° grid and with a monthly temporal resolution over 1998–2015. It is referred hereafter to as $\mathrm{Chl}_{\text{OC-CCI}}$.

Predictors have been extracted over the same time period than $\mathrm{Chl}_{\text{OC-CCI}}$. Those with a higher resolution than 1° and a month have been averaged to match the Chl grid. The 2° × 2° bins of the short-wave radiation product have been divided to match the Chl grid.

The Multivariate El Niño Southern Oscillation Index (MEI) has been provided by the National Oceanic and Atmospheric Administration (NOAA) website [34].

2.2. Machine Learning Models

2.2.1. Support Vector Regression (SVR)

Support vector machine is a kernel-based supervised learning method [35] developed for classification purposes in the early 1990s and then extended for regression by [36]. The basic idea behind SVR is to map the variables into a new space, possibly in a non-linear way using the so-called kernel function, so that the regression task hopefully becomes linear in this space. Because SVR can efficiently capture complex non-linear relationships, it has been used in a variety of fields, and more specifically for oceanographic, meteorological and climate impact studies [37–39], as well as in marine bio-optics [40–42]. Considering a Gaussian kernel, SVR only involves the selection of two hyperparameters: the penalty parameter C of the error term and the kernel band with gamma. Following [16], (i) these two parameters have been set to 2 and 0.3, respectively; and (ii) the SVR was trained on only 9% of the database (randomly selected) due to computational limitations. The SVR is set up with python and the Scikit-learn library. Reconstructed Chl is hereafter referred to as $\mathrm{Chl}_{\text{SVR}}$.

2.2.2. Neural Networks

Deep learning models and neural networks (NNs) are at the core of the state-of-the-art in machine learning and artificial intelligence for a large range of applications [43]. NNs are particularly appealing due to their capacity to learn complex relationships from raw data better than other models, when data are abundant. Though the concept of NNs has been around for a long time, these state-of-the-art approaches recently obtained impressive results for many supervised or unsupervised learning problems thanks to the availability of very large datasets, and the increase in computational power in the last few decades [44]. NNs learn complex patterns through the composition of simple elementary operations forming a global highly-nonlinear input/output relationships, whose parameters can be efficiently trained using the back-propagation algorithm. These recent advances motivate an increasing number of applications in spatial oceanography, (e.g., [45,46]). In this context, physics-informed and theory-guided NNs [47] are of key interest as new means to exploit the computational and learning efficiency of NNs, while exploiting prior knowledge and making easier model interpretation. Here, we follow such an approach with a MLP architecture, which uses the same physical features and geospatial information as inputs as the SVR.

NN frameworks exploit mini-batch gradient descent schemes during the training phase. This can allow us to exploit the entire dataset contrary to the SVR model. Indeed, SVR is known not to scale up well with large training datasets. The considered MLP architecture exploits LeakyReLU activation functions after each dense layer (Figure S1). Dropout layers [48] are added to the last 3 layers to reduce overfitting. Since we are dealing with a regression task, the last layer is linear. Thanks to its benefit of penalizing large errors, MSE is used as the loss function for the MLP. The MLP is set up with python and the Keras library. Configurations details are provided in Table S1.

As for the SVR, we first train a MLP on 9% of the training dataset, randomly selected. The reconstructed Chl is hereafter referred to as $\mathrm{Chl}_{\text{MLP-9\%}}$. The aim is to provide a first consistent comparison with

Chl_{SVR} using the same training setting. Then, we take advantage of the MLP ability to be trained on larger database and a second MLP is trained on 80% of the dataset. Reconstructed Chl is referred to as Chl_{MLP}. The two learning curves are provided in Figure S2.

2.2.3. Data Preprocessing and Procedure

Predictors and log(Chl) are normalized by removing their respective average and dividing them by their standard deviations. The SVR and the two MLP are trained from 1998 (the first complete year of the satellite $\text{Chl}_{\text{OC-CCI}}$ time-series) to 2015 between 50° S and 50° N (Step 1 in Figure 1). Thus, the resulting SVR and NN schemes are applied on the physical predictors over 1998–2015, and the annual means and standard deviations initially removed are applied to perform the back transformation and reconstruct Chl values, namely either Chl_{SVR}, $\text{Chl}_{\text{MLP-9\%}}$ and Chl_{MLP} outputs (Step 2 in Figure 1). These three Chl reconstructed whole datasets are then compared to $\text{Chl}_{\text{OC-CCI}}$ to evaluate their skills in reconstructing satellite observations (Step 3 in Figure 1).

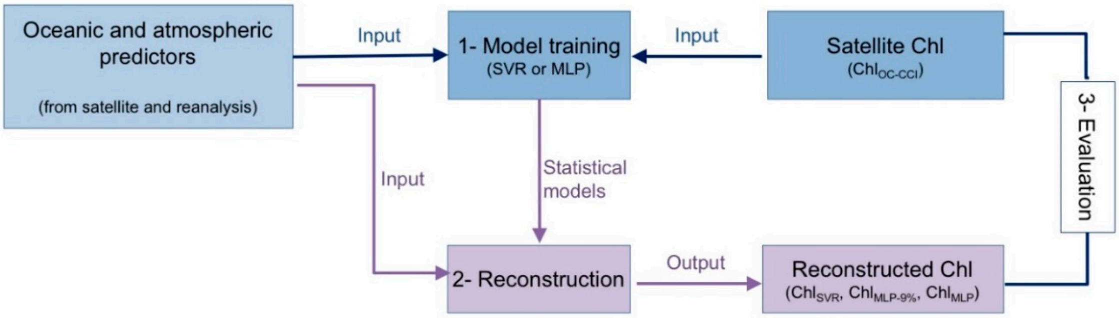
Oceanic and atmospheric predictors
(from satellite and reanalysis)
Input
1- Model training
(SVR or MLP)
Input
Satellite Chl
($\text{Chl}_{\text{OC-CCI}}$)
Input
Statistical models
3- Evaluation
2- Reconstruction
Output
Reconstructed Chl
(Chl_{SVR}, $\text{Chl}_{\text{MLP-9\%}}$, Chl_{MLP})

Figure 1. Three-step procedure to train the machine learning models, reconstruct surface Chlorophyll-a concentration (Chl) and evaluate the statistical model skills over 1998–2015.

2.3. *Statistical Diagnostics and Empirical Orthogonal Functions*

First, scatter plots are performed to compare satellite vs. reconstructed Chl for the Atlantic, Pacific and Indian Oceans between 50° S and 50° N. Root mean square error (RMSE) is derived at basin-scale as $RMSE = \sqrt{\sum \frac{(Chl_{reconstructed} - Chl_{OC-CCI})^2}{N}}$, with N the sample number. Pearson correlation and normalized RMSE (NRMSE) are also derived, with NRMSE = $\frac{RMSE}{<Chl_{OC-CCI}>}$ and $< Chl_{OC-CCI} >$ the mean $\text{Chl}_{\text{OC-CCI}}$ value.

Empirical orthogonal function (EOF) analysis is performed to investigate the model's ability to reconstruct Chl seasonal and interannual variability. First, centered seasonal and interannual $\text{Chl}_{\text{OC-CCI}}$ are obtained by removing their annual and monthly means over 1998–2015, respectively, and by dividing them by their standard deviations. Then, the so-called reference EOF analysis is performed on these centered seasonal and interannual $\text{Chl}_{\text{OC-CCI}}$ anomalies to avoid an overly dominant contribution of high values on the analysis [49]. Thus, Chl_{SVR} and Chl_{MLP} outputs are projected onto the seasonal and interannual $\text{Chl}_{\text{OC-CCI}}$ spatial patterns to obtain their associated time components (i.e., principal component-PC). Seasonal and interannual PCs for each dataset are then compared.

3. Results

3.1. *Statistical Performances*

A first evaluation of the Chl_{SVR} vs. $\text{Chl}_{\text{OC-CCI}}$ is provided over 1998–2015 at basin scales and for the whole dataset (Figure 2, upper row). Determination coefficients between both datasets are below 0.5 and even get down to 0.26 in the Indian Ocean, while RMSE is about 0.6 in the three basins. The MLP trained on the same amount of data than the SVR is more skillful than the SVR to reconstruct $\text{Chl}_{\text{OC-CCI}}$ (Figure 2, middle row). The regression lines between the log of $\text{Chl}_{\text{MLP-9\%}}$ vs. $\text{Chl}_{\text{OC-CCI}}$ are closer to the 1:1 line for each oceanic basin with determination coefficients higher than 0.63 and up to 0.73 in the Atlantic Ocean. Increasing from 9% to a usual 80% the MLP training dataset further increase the skills of the NN approach to reconstruct Chl$_{\text{OC-CCI}}$ with a relative gain of about 10% (Figure 2, lower row). Nevertheless, Chl$_{\text{MLP}}$ still underestimates Chl$_{\text{OC-CCI}}$, more specifically in the Pacific. Some of these differences may be related to changes in Chl, which may due for instance to photo acclimation (e.g., [50,51]) or by other components that are not Chl such as suspended particulate matter (SPM) or colored dissolved organic matter (CDOM; [52]). Interestingly, the use of a MLP not only removes computational restrictions imposed by the SVR (i.e., size of the training samples), but it also appears to be more efficient in reconstructing surface Chl from oceanic and atmospheric variables. Thereafter, Chl$_{\text{OC-CCI}}$ and Chl$_{\text{SVR}}$ are compared to the best NN product, Chl$_{\text{MLP}}$ (i.e., trained on 80% of the dataset).

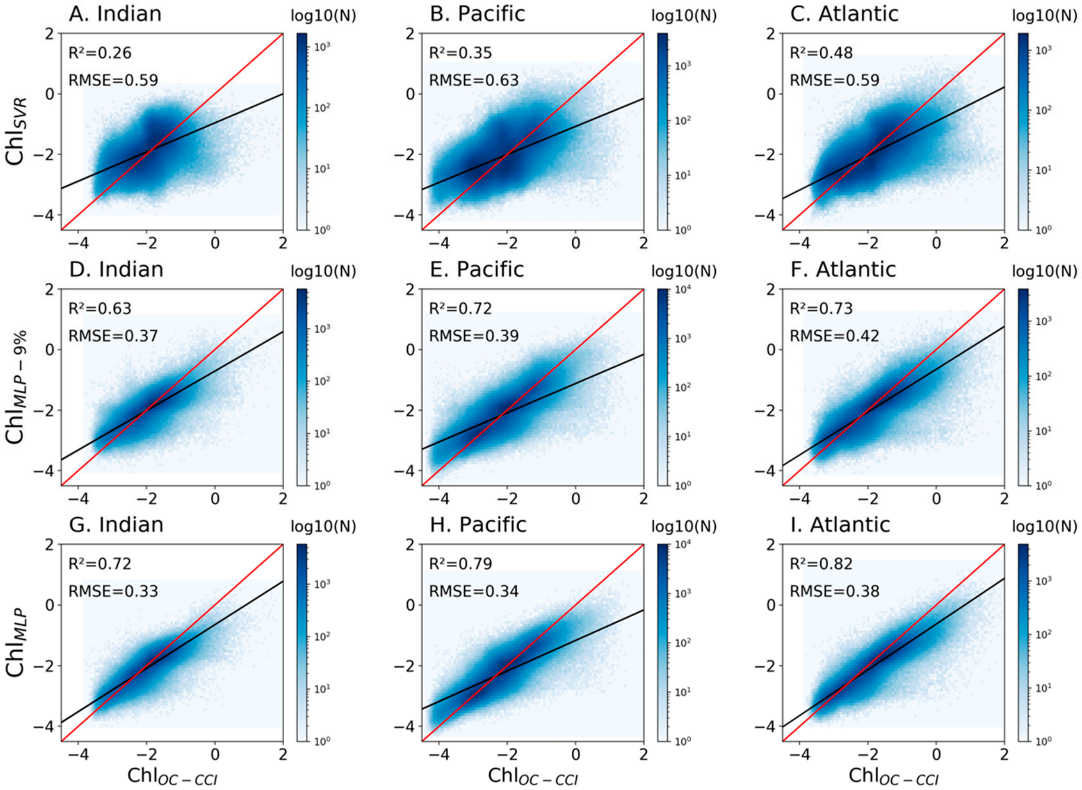

Figure 2. Scatter plots of log of Chl$_{\text{OC-CCI}}$ vs. (**A**–**C**) Chl$_{\text{SVR}}$, (**D**–**F**) Chl$_{\text{MLP-9\%}}$ and (**G**–**I**) Chl$_{\text{MLP}}$ trained on 80% of the dataset, for each oceanic basin between 50° S and 50° N and over 1998–2015. The Chl$_{\text{OC-CCI}}$ vs. reconstructed Chl regression lines are plotted in black and the 1:1 regression lines are plotted in red. The figure is color-coded according to the density of observations.

Consistently with the scatterplots, Chl$_{\text{OC-CCI}}$ correlations with Chl$_{\text{SVR}}$ are significantly lower than with Chl$_{\text{MLP}}$ (Figure 3A,C; r = 0.17 vs. 0.6 on a global scale, respectively). Chl$_{\text{OC-CCI}}$ - Chl$_{\text{SVR}}$ correlations are higher than 0.7 ($p < 0.001$) over limited regions such as the Atlantic, Indian, and Pacific subtropical areas (Figure 3A). The MLP allows a significant improvement in the correlation with Chl$_{\text{OC-CCI}}$ with values higher than 0.75 over most of the global ocean (Figure 3C). Areas of high and low NRMSE are similarly distributed for Chl$_{\text{SVR}}$ and Chl$_{\text{MLP}}$ (Figure 3B,D). For instance, in both cases NRMSE is higher at the highest latitudes and in the tropical north-western and south-eastern Pacific. Although the MLP reduces these NRMSE by 50% compared to the SVR, biases in reference to Chl$_{\text{OC-CCI}}$ still remain in these regions. High NRMSE can reflect the influence of other components than phytoplankton biomass on the Chl signal as mentioned above, or/and the impact of other predictors not considered to train the neural network models. This is, for instance, suggested from the Amazon plume where the Chl signal is known to be influenced by river flow and may thus rather be associated with CDOM or SPM than phytoplankton biomass [53].

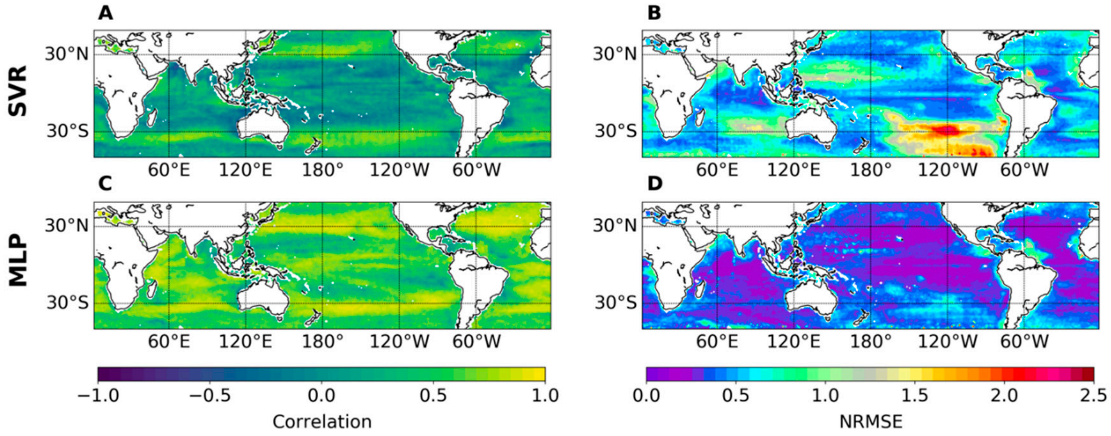

Figure 3. (**A**,**C**) Correlation and (**B**,**D**) NRMSE of $Chl_{OC\text{-}CCI}$ vs. (up) Chl_{SVR} and (bottom) Chl_{MLP} over 1998–2015.

3.2. Seasonal to Interannual Variability and Trends

EOF analysis provides complementary insights into the MLP and SVR ability in reconstructing Chl spatio-temporal variability. Seasonal and interannual $Chl_{OC\text{-}CCI}$ spatial variability are illustrated through their respective EOF first modes with a total variance of 27.6% and 12.6%, respectively (Figure 4A,C). The well-known Chl seasonal patterns are highlighted with a variability out of phase between the northern vs. southern hemisphere due to the reversal of the season order (Figure 4A,B). It is also out of phase between high latitudes (light limited) vs. low and mid-latitudes (rather nutrient limited). Chl_{MLP} and Chl_{SVR} are then projected on this reference $Chl_{OC\text{-}CCI}$ spatial seasonal pattern. Correlations between $Chl_{OC\text{-}CCI}$ and Chl_{SVR} or Chl_{MLP} PCs are of 0.64 and 0.99, respectively (Figure 4B). Chl_{SVR} PC shows a fictive double peak in the seasonal cycle and a largely underestimated amplitude. Despite the $Chl_{OC\text{-}CCI}$ - Chl_{MLP} high correlation, the amplitude of Chl_{MLP} seasonal variability is also underestimated.

The first EOF mode performed on interannual $Chl_{OC\text{-}CCI}$ illustrates the largely reported impact of El Nino Southern Oscillation (ENSO) [4,17,54–56] (Figure 4C). Here again, the MLP results in a significant improvement compared with the SVR to reconstruct $Chl_{OC\text{-}CCI}$ variability, with its first PC correlation with $Chl_{OC\text{-}CCI}$ of 0.95 vs. 0.63 for Chl_{SVR} (Figure 4D).

Over the last 18 years, observed $Chl_{OC\text{-}CCI}$ trends have increased over most of the global ocean (Figure 5A). Regionally, some decreases are observed such as in the Indian Ocean, the equatorial Pacific and the Atlantic and Pacific oligotrophic subtropical gyres. While most of these trends are captured by Chl_{SVR}, inverse trends occur in the South Indian and Atlantic Oceans (Figure 5B). In addition, the trend estimation for Chl_{SVR} reveals an unrealistic high-frequency pattern, which may relate to the support of the Gaussian kernels implemented by the SVR model. On its side, Chl_{MLP} better reproduces $Chl_{OC\text{-}CCI}$ trends in terms of spatial distribution, although their amplitude remain underestimated, especially in the Indian Ocean (Figure 5C).

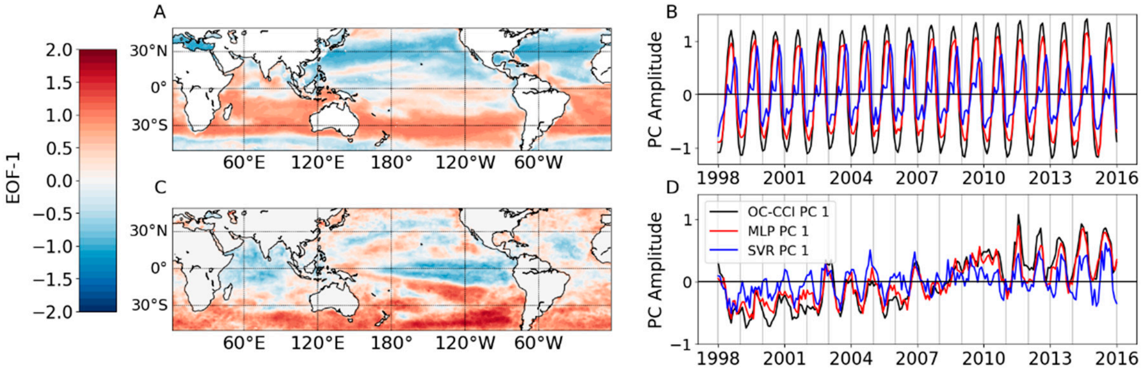

Figure 4. First mode of the $\text{Chl}_{\text{OC-CCI}}$ (**A**) seasonal and (**C**) interannual EOFs over 1998–2015, and their associated principal components (**B**,**D**), respectively. Chl_{MLP} and Chl_{SVR} PCs obtained from the $\text{Chl}_{\text{OC-CCI}}$ EOF projections are reported in (**B**,**D**).

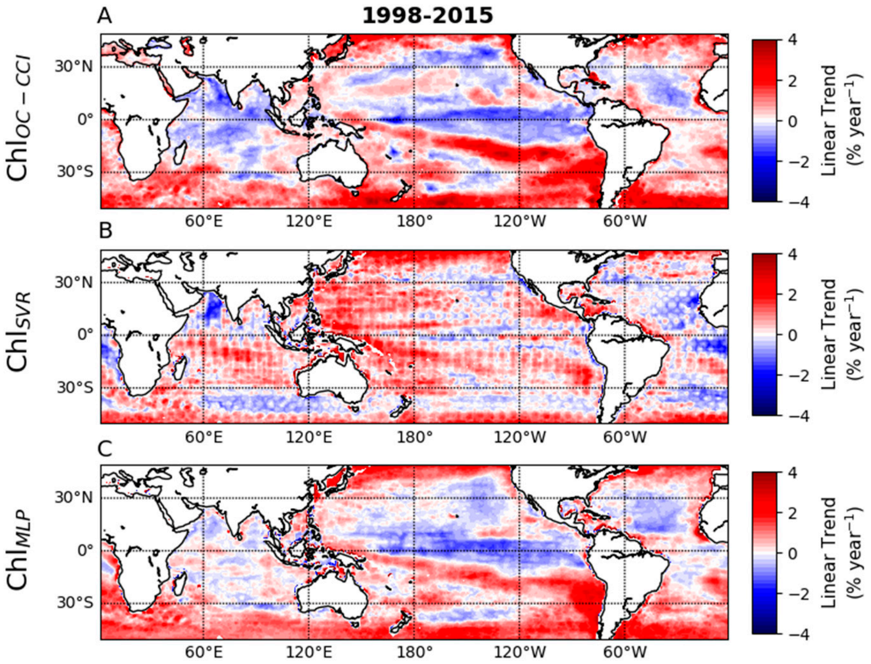

Figure 5. Linear log of (Chl) trend (in % year^{-1}) calculated over 1998–2015 from the monthly (**A**) $\text{Chl}_{\text{OC-CCI}}$, (**B**) ChlSVR, (**C**) ChlMLP.

A similar conclusion can be done when considering Chl trends in oligotrophic gyres (with surface Chl < 0.07 mg.m^{-3} as in [18]) over the last 15 years. Such trends can be weak and not sufficiently well resolved by the MLP (nor the SVR) in terms of sign or amplitude (Figure 6). While it is crucial to improve the reconstruction of the signal amplitude on a global scale, it is particularly appealing in those oceanic deserts which would tend to expand in the context of climate warming and increasing stratification [18,57,58]. As expected, the spread of these oligotrophic areas is different here from $\text{Chl}_{\text{OC-CCI}}$ than those observed by [18] based on a single-sensor Chl product over 1998–2006 [59]. Thus, training deep learning schemes on Chl products from both single and multi-sensors could provide a more synoptic view on the impact and interpretation of reconstructed Chl long time-series.

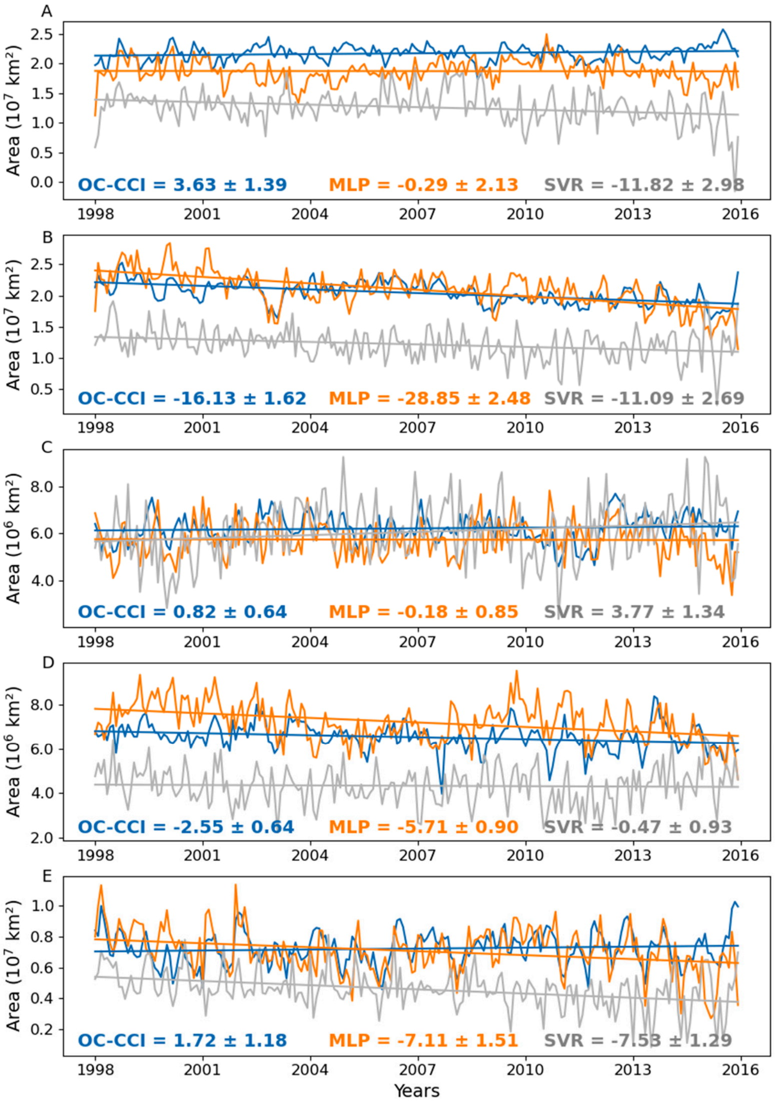

Figure 6. Time series of the monthly mean area (km^2) with surface Chl less than or equal to 0.07 mg.m^{-3} between 5° and 45° N/° S latitude with the seasonal cycle removed in the (**A**) North Pacific, (**B**) South Pacific, (**C**) North Atlantic, (**D**) South Atlantic and (**E**) Indian Oceans. Chl$_{\text{OC-CCI}}$ is in blue, Chl$_{\text{SVR}}$ in grey and Chl$_{\text{MLP}}$ in orange. Straight lines are the linear trends calculated as in Figure 5. The average trends and standard deviation are also indicated.

Finally, a first attempt to investigate predictors' relative importance is performed with the MLP approach and the mean decrease accuracy method [60]. As expected, the SST and short-wave radiations are the most important predictors (Table 2). Interestingly, the zonal surface wind stress component seems of particular importance while sea level anomaly is far behind. Indeed, redundant indirect information's about the ocean circulation can be derived from those predictors, with potentially more information "embedded" within the wind stress that may also be linked to a meaningful parameter for phytoplankton growth: the mixed layer depth. Spatial coordinates (latitude and longitude) are less important for the reconstructions suggesting that the MLP can extract geospatially-dependent features from other predictors than the coordinates themselves. Thus, a complementary run was performed, still using 80% of the dataset to train the MLP but based only on the five main predictors with a relative importance higher than 0.1 in Table 2. As expected, removing spatial information's from the MLP training still allows to reconstruct realistic Chl (unlike the SVR, which produces unrealistic concentrations). Interestingly, although the averaged relative importance of seven of the eight removed predictors are lower than 0.033, the determination coefficients between this specific

Chl_{MLP} vs. $\text{Chl}_{\text{OC-CCI}}$ decrease by more than 0.2 while the RMSE increase to more than 0.12 in each oceanic basin when compared to the MLP using all the predictors (Figure S3 vs. Figure 2). This result illustrates that the use of these predictors, apparently weakly significant in average, does not change significantly the Chl geographical distribution pattern, but can modulate its regional intensity (Figure S4). The impact of the different predictors on Chl reconstruction according to the oceanic regions and/or the climate cycles should therefore be specifically considered and investigated in a dedicated study, once a deep learning scheme will have been considered as sufficiently satisfying.

Table 2. Multi-layer perceptron (MLP) predictor's relative importance.

Weight	Predictors
0.6691 +/− 0.0021	SST
0.2672 +/− 0.001	Short-wave radiations
0.2316 +/− 0.0013	Zonal surface wind
0.1526 +/− 0.0008	cos(month)
0.1449 +/− 0.0004	Meridional surface wind
0.0641 +/− 0.0003	sin(month)
0.0328 +/− 0.0004	Zonal surface current
0.0328 +/− 0.0004	Sea level anomalies
0.0260 +/− 0.0003	Year
0.0180 +/− 0.0002	Meridional surface current
0.0002 +/− 0.0001	lat
0.0002 +/− 0.0001	cos(lon)
0.0002 +/− 0.0001	sin(lon)

4. Discussion

To our knowledge, the only study that has been performed to reconstruct satellite surface Chl on a global scale used a SVR approach [16]. In this former study, we also used $\text{Chl}_{\text{OC-CCI}}$ and the same physical predictors than in the present study, but the predictors were originating from a numerical model (vs. satellite observations and numerical atmospheric reanalysis here). The SVR accurately reproduced most aspects of the satellite Chl variability (although underestimated by half) and spatial patterns. Here, we show that the SVR, trained on satellite data and atmospheric reanalysis, also encounters difficulties to well reproduce the Chl signal. The MLP demonstrates the ability of deep learning schemes to reproduce satellite Chl with far better skill than the SVR, not only to capture the general patterns of Chl seasonal and interannual signals and trends, but also their amplitude. Neither the training of the MLP nor that of the SVR involve time information, as the training loss only involves a grid point-wise reconstruction error criterion. Thus, our results support the greater ability of the MLP to generalize time patterns than the SVR and the relevance of neural network approaches compared with kernel methods.

However, further efforts still remain to alleviate the issue of Chl underestimation. One plausible hypothesis would be that complementary predictors may provide additional insights to this issue. For instance, precipitation is among a key driver of coastal run-off and river discharge could be considered. Thus, applying similar learning-based schemes to CDOM and SPM, especially jointly to Chl, would be note-worthy. Such combined approaches could help improving both reconstruction and interpretation of Chl in regions where satellite Chl products from case 1 waters may reflect changes from other components than phytoplankton biomass. Particulate backscattering (as a proxy of the Particulate Organic Carbon) retrieved from satellite also deserves to be considered in addition to Chl in the training/reconstruction process. Indeed, it would allow us to investigate the extent to which the Chl variability reflects changes in phytoplankton biomass vs. cellular changes in response to light [51,61,62], which is of particular importance in oligotrophic gyres.

If, in this study, we demonstrate the better potential of NNs to accurately represent Chl seasonal and interannual signals, compared to the SVR, so far, only a MLP was used. MLP is known to not explicitly consider the spatial and temporal correlations in the dataset. Specific architectures to handle spatially or temporally structured data, i.e., convolutional neural networks and recurrent neural networks (such as long short-term memory networks) are currently under investigation and are expected to further improve the Chl reconstruction performance, in particular for the Chl seasonal cycle amplitude. Interestingly, such NN architectures would provide additional insights on the physical variables of interest through sensitivity tests, which could drive the reconstruction of Chl beyond the set of predictors considered in this study. Recent advances in deep learning for irregularly-sampled datasets also suggest that future work could learn such NN representations from datasets involving missing data patterns, which may be of key interest for Chl patterns in some regions [63,64].

5. Conclusions

The present study investigates two statistical approaches to derive from satellite observations the Chl seasonal to interannual variability and trends, as well as their potential in reconstructing biological past long-term time-series. The MLP is more skillful than the SVR to capture both the spatio-temporal patterns and amplitude of $\mathrm{Chl_{OC\text{-}CCI}}$ on a global scale over 1998–2015. $\mathrm{Chl_{MLP}}$ and $\mathrm{Chl_{OC\text{-}CCI}}$ seasonal and interannual first mode of variability are highly correlated ($\mathrm{r\text{-}_{PCs}}$ = 0.99 and 0.95, $p < 0.001$), suggesting that the MLP is reliable to reproduce $\mathrm{Chl_{OC\text{-}CCI}}$, at least over the last 18 years. However, some underestimation in $\mathrm{Chl_{MLP}}$ amplitudes as well as regional bias need to be fixed in future studies and the predictors' importance in Chl reconstruction deserve to be more deeply investigated.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-4292/12/24/4156/s1, Figure S1: MLP architecture. Table 1: MLP configuration. Figure S2: Learning curves of $\mathrm{Chl_{MLP\text{-}9\%}}$ and $\mathrm{Chl_{MLP}}$. Figure S3: Scatter plots of $\mathrm{Chl_{MLP}}$ trained only for predictors with a relative importance higher than 0.1 in Table 2. Figure S4: Correlation and NRMSE of $\mathrm{Chl_{OC\text{-}CCI}}$ vs. $\mathrm{Chl_{MLP}}$ trained only for predictors with a relative importance higher than 0.1 in Table 2.

Author Contributions: Conceptualization, E.M., A.B., T.G., L.D., J.R., P.T., G.M. and R.F.; data creation, A.B. and J.R.; methodology, A.B., L.D., R.F., P.T., E.M., T.G. and G.M.; formal analysis, A.B. and J.R.; writing—original draft preparation, E.M.; writing—review and editing, T.G., L.D., R.F., P.T., J.R., A.B. and G.M.; supervision, E.M., T.G., L.D. and R.F.; funding acquisition, E.M. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for the internship of A. Brini and GPU was provided by the CNES/TOSCA call (Grant n°160515/00) PhytoDeV project. The J. Roussillon PhD grant was funded by the ISblue project, Interdisciplinary graduate school for the blue planet (ANR-17-EURE-0015) and co-funded by a grant from the French government under the program "Investissements d'Avenir".

Acknowledgments: We would like to thank the three anonymous reviewers whose comments have helped to improve the manuscript.

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

References

1. Sabine, C.L.; Feely, R.A.; Gruber, N.; Key, R.M.; Lee, K.; Bullister, J.L.; Wanninkhof, R.; Wong, C.S.; Wallace, D.W.R.; Tilbrook, B.; et al. The oceanic sink for anthropogenic CO_2. *Science* **2004**, *305*, 367–371. [CrossRef]
2. Falkowski, P. Ocean science: The power of plankton. *Nature* **2012**, *483*, S17–S20. [CrossRef]
3. Longhurst, A.; Sathyendranath, S.; Platt, T.; Caverhill, C. An estimate of global primary production in the ocean from satellite radiometer data. *J. Plankton Res.* **1995**, *17*, 1245–1271. [CrossRef]
4. Messié, M.; Chavez, F.P. A global analysis of ENSO synchrony: The oceans' biological response to physical forcing. *J. Geophys. Res. Oceans* **2012**, *117*, C09001. [CrossRef]
5. Dutkiewicz, S.; Follows, M.; Marshall, J.; Gregg, W.W. Interannual variability of phytoplankton abundances in the North Atlantic. *Deep Sea Res. II Top. Stud. Oceanogr.* **2001**, *48*, 2323–2344. [CrossRef]

6. Aumont, O.; Ethé, C.; Tagliabue, A.; Bopp, L.; Gehlen, M. PISCESv2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* **2015**, *8*, 2465–2513. [CrossRef]
7. Henson, S.A.; Dunne, J.P.; Sarmiento, J.L. Decadal variability in North Atlantic phytoplankton blooms. *J. Geophys. Res. Ocean.* **2009**, *114*, C04013. [CrossRef]
8. Henson, S.A.; Raitsos, D.; Dunne, J.P.; McQuatters-Gollop, A. Decadal variability in biogeochemical models: Comparison with a 50-year ocean colour dataset. *Geophys. Res. Lett.* **2009**, *36*, L21061. [CrossRef]
9. Patara, L.; Visbeck, M.; Masina, S.; Krahmann, G.; Vichi, M. Marine biogeochemical responses to the North Atlantic Oscillation in a coupled climate model. *J. Geophys. Res. Ocean.* **2011**, *116*, C07023. [CrossRef]
10. Trenberth, K.E.; Fasullo, J.T. An apparent hiatus in global warming? *Earth's Future* **2013**, *1*, 19–32. [CrossRef]
11. Wilson, C.; Adamec, D. A global view of bio-physical coupling from SeaWiFS and TOPEX satellite data, 1997–2001. *Geophys. Res. Lett.* **2002**, *29*, 1257. [CrossRef]
12. Wilson, C.; Coles, V.J. Global climatological relationships between satellite biological and physical observations and upper ocean properties. *J. Geophys. Res. Ocean.* **2005**, *110*, C10001. [CrossRef]
13. Kahru, M.; Gille, S.T.; Murtugudde, R.; Strutton, P.G.; Manzano-Sarabia, M.; Wang, H.; Mitchell, B.G. Global correlations between winds and ocean chlorophyll. *J. Geophys. Res. Ocean.* **2010**, *115*, C12040. [CrossRef]
14. Messié, M.; Chavez, F.P. Seasonal regulation of primary production in eastern boundary upwelling systems. *Prog. Oceanogr.* **2015**, *134*, 1–18. [CrossRef]
15. Uz, S.S.; Busalacchi, A.J.; Smith, T.M.; Evans, M.N.; Brown, C.W.; Hackert, E. Interannual and decadal variability in tropical pacific chlorophyll from a statistical reconstruction: 1958–2008. *J. Clim.* **2017**, *30*, 7293–7315. [CrossRef]
16. Martinez, E.; Gorgues, T.; Lengaigne, M.; Fontana, C.; Sauzède, R.; Menkes, C.; Uitz, J.; Di Lorenzo, E.; Fablet, R. Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach. *Front. Mar. Sci.* **2020**, *7*, 464. [CrossRef]
17. Behrenfeld, M.J.; O'Malley, R.T.; Siegel, D.A.; McClain, C.R.; Sarmiento, J.L.; Feldman, G.C.; Milligan, A.J.; Falkowski, P.G.; Letelier, R.M.; Boss, E.S. Climate-driven trends in contemporary ocean productivity. *Nature* **2006**, *444*, 752–755. [CrossRef]
18. Polovina, J.J.; Howell, E.A.; Abecassis, M. Ocean's leastproductive waters are expanding. *Geophys. Res. Lett.* **2008**, *35*, L03618. [CrossRef]
19. Martinez, E.; Antoine, D.; D'Ortenzio, F.; Gentili, B. Climate-driven basin-scale decadal oscillations of oceanic phytoplankton. *Science* **2009**, *36*, 1253–1256. [CrossRef]
20. Thomas, A.C.; Strub, P.T.; Weatherbee, R.A.; James, C. Satellite views of Pacific chlorophyll variability: Comparisons to physical variability, local versus nonlocal influences and links to climate indices. *Deep-Sea Res. II Top. Stud. Oceanogr.* **2012**, *77*, 99–116. [CrossRef]
21. Lewandowska, A.M.; Hillebrand, H.; Lengfellner, K.; Sommer, U. Temperature effects on phytoplankton diversity—The zooplankton link. *J. Sea Res.* **2014**, *85*, 359–364. [CrossRef]
22. Available online: http://iridl.ldeo.columbia.edu/ (accessed on 18 December 2020).
23. Wilson, C.; Adamec, D. Correlations between surface chlorophyll and sea surface height in the tropical Pacific during the 1997-1999 El Nino-Southern event. *J. Geophys. Res. Ocean.* **2001**, *106*, 31175–31188. [CrossRef]
24. Radenac, M.H.; Léger, F.; Singh, A.; Delcroix, T. Sea surface chlorophyll signature in the tropical Pacific during eastern and central Pacific ENSO events. *J. Geophys. Res. Ocean.* **2012**, *117*, C04007. [CrossRef]
25. Available online: https://resources.marine.copernicus.eu/?option=com_csw&task=results (accessed on 18 December 2020).
26. Martinez, E.; Antoine, D.; D'Ortenzio, F.; de Boyer Montégut, C. Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s. *J. Geophys. Res. Ocean.* **2011**, *116*, C11029. [CrossRef]
27. Available online: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim (accessed on 18 December 2020).
28. Radenac, M.H.; Messié, M.; Léger, F.; Bosc, C. A very oligotrophic zone observed from space in the equatorial Pacific warm pool. *Remote Sens. Environ.* **2013**, *134*, 224–233. [CrossRef]
29. Available online: https://podaac.jpl.nasa.gov/dataset/OSCAR_L4_OC_third-deg (accessed on 18 December 2020).
30. Available online: https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.surfaceflux.html (accessed on 18 December 2020).

31. Sauzède, R.; Claustre, H.; Jamet, C.; Uitz, J.; Ras, J.; Mignot, A.; D'Ortenzio, F. Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications. *J. Geophys. Res. Ocean.* **2015**, *120*, 451–470. [CrossRef]
32. Available online: https://www.oceancolour.org (accessed on 18 December 2020).
33. Belo Couto, A.; Brotas, V.; Mélin, F.; Groom, S.; Sathyendranath, S. Inter-comparison of OC-CCI chlorophyll—A estimates with precursor data sets. *Int. J. Remote Sens.* **2016**, *37*, 4337–4355. [CrossRef]
34. Available online: www.esrl.noaa.gov/psd (accessed on 18 December 2020).
35. Vapnik, V. Statistics for engineering and information science. In *The Nature of Statistical Learning Theory*; Jordan, M.J., Lawless, J.F., Lauritzen, S.L., Nair, V., Eds.; Springer: New York, NY, USA, 2000.
36. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
37. Descloux, E.; Mangeas, M.; Menkes, C.E.; Lengaigne, M.; Leroy, A.; Tehei, T.; Guillaumot, L.; Teurlai, M.; Gourinat, A.-C.; Benzler, J.; et al. Climate-based models for understanding and forecasting dengue epidemics. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1470. [CrossRef]
38. Elbisy, M.S. Sea wave parameters prediction by support vector machine using a genetic algorithm. *J. Coast. Res.* **2015**, *31*, 892–899. [CrossRef]
39. Neetu, S.; Lengaigne, M.; Vialard, J.; Mangeas, M.; Menkes, C.; Suresh, I.; Leloup, J.; Knaff, J. Quantifying the benefits of non-linear methods for global statistical hindcasts of tropical cyclones intensity. *Mon. Weather Rev.* **2020**, *35*, 807–820. [CrossRef]
40. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.K.; Ha, S. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *Gisci. Remote Sens.* **2014**, *51*, 158–174. [CrossRef]
41. Hu, S.; Liu, H.; Zhao, W.; Shi, T.; Hu, Z.; Li, Q.; Wu, G. Comparison of machine learning techniques in inferring phytoplankton size classes. *Remote Sens.* **2018**, *10*, 191. [CrossRef]
42. Blix, K.; Eltoft, T. Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sens.* **2018**, *10*, 775. [CrossRef]
43. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
44. Ahmad, H. Machine learning applications in oceanography. *Aquat. Res.* **2019**, *2*, 161–169. [CrossRef]
45. Sammartino, M.; Marullo, S.; Santoleri, R.; Scardi, M. Modelling the vertical distribution of phytoplankton biomass in the Mediterranean Sea from satellite data: A neural network approach. *Remote Sens.* **2018**, *10*, 1666. [CrossRef]
46. Wang, C.; Tandeo, P.; Mouche, A.; Stopa, J.E.; Gressani, V.; Longepe, N.; VanDeMark, D.; Foster, R.C.; Chapron, B. Classification of the global Sentinel-1 SAR vignettes for ocean surface process studies. *Remote Sens. Environ.* **2019**, *234*, 111457. [CrossRef]
47. Karpatne, A.; Atluri, G.; Faghmous, J.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-guided data science: A new paradigm for scientific discovery. *arXiv* **2016**, arXiv:1612.08544. [CrossRef]
48. Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Srivastva, N. System and Method for Addressing Overfitting in a Neural Network. US Patent 9,406,017, 2 August 2016.
49. Emery, W.; Thomson, R. *Data Analysis in Physical Oceanography*; Pergamon: New York, NY, USA, 1997; p. 634.
50. Laws, E.A.; Bannister, T.T. Nutrient-and light-limited growth of Thalassiosira fluviatilis in continuous culture, with implications for phytoplankton growth in the ocean. *Limnol. Oceanogr.* **1980**, *25*, 457–473. [CrossRef]
51. Behrenfeld, M.J.; O'Malley, R.T.; Boss, E.S.; Westberry, T.K.; Graff, J.R.; Halsey, K.H.; Milligan, A.J.; Siegel, D.A.; Brown, M.B. Revaluating ocean warming impacts on global phytoplankton. *Nat. Clim. Chang.* **2015**, *6*, 323–330. [CrossRef]
52. Morel, A.; Gentili, B. The dissolved yellow substance and the shades of blue in the Mediterranean Sea. *Biogeosciences* **2009**, *6*, 2625–2636. [CrossRef]
53. Del Vecchio, R.; Subramaniam, A. Influence of the Amazon River on the surface optical properties of the western tropical North Atlantic Ocean. *J. Geophys. Res. Ocean.* **2004**, *109*, C11. [CrossRef]
54. Behrenfeld, M.J.; Randerson, J.T.; McClain, C.R.; Feldman, G.C.; Los, S.O.; Tucker, C.J.; Falkowski, P.G.; Field, C.B.; Frouin, R.; Esaias, W.E.; et al. Biospheric primary production during an ENSO transition. *Science* **2001**, *291*, 2594–2597. [CrossRef] [PubMed]
55. Yoder, J.A.; Kennelly, M.A. Seasonal and ENSO variability in global ocean phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements. *Glob. Biogeochem. Cycles* **2003**, *17*, 1112. [CrossRef]

56. Chavez, F.P.; Messié, M.; Pennington, J.T. Marine primary production in relation to climate variability and change. *Annu. Rev. Mar. Sci.* **2011**, *3*, 227–260. [CrossRef] [PubMed]
57. Doney, S.C. Plankton in a warmer world. *Nature* **2006**, *444*, 695–696. [CrossRef] [PubMed]
58. Irwin, A.J.; Oliver, M.J. Are ocean deserts getting larger? *Geophys. Res. Lett.* **2009**, *36*, L18609. [CrossRef]
59. Mélin, F.; Vantrepotte, V.; Chuprin, A.; Grant, M.; Jackson, T.; Sathyendranath, S. Assessing the fitness-for-purpose of satellite multi-mission ocean color climate data records: A protocol applied to OC-CCI chlorophyll-a data. *Remote Sens. Environ.* **2017**, *203*, 139–151. [CrossRef]
60. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
61. Siegel, D.A.; Maritorena, S.; Nelson, N.B.; Behrenfeld, M.J. Independence and interdependencies among global ocean color properties: Reassessing the bio-optical assumption. *J. Geophys. Res. Ocean.* **2005**, *110*, C07011. [CrossRef]
62. Westberry, T.; Behrenfeld, M.J.; Siegel, D.A.; Boss, E. Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Glob. Biogeochem. Cycles* **2008**, *22*, GB2024. [CrossRef]
63. Beauchamp, M.; Fablet, R.; Ubelmann, C.; Ballarotta, M.; Chapron, B. Intercomparison of data-driven and learning-based interpolations of along-track Nadir and wide-swath Swot altimetry observations. *Remote Sens.* **2020**, *12*, 3806. [CrossRef]
64. Nguyen, D.; Ouala, S.; Drumetz, L.; Fablet, R. Variational Deep Learning for the Identification and Reconstruction of Chaotic and Stochastic Dynamical Systems from Noisy and Partial Observations. *arXiv* **2020**, arXiv:2009.02296.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.