The Synergy of Data From Profiling Floats, Machine Learning and Numerical Modeling: Case of the Black Sea Euphotic Zone

E. V. Stanev[1,2,3] , K. Wahle[1] , and J. Staneva[1]

[1]Institute of Coastal Systems—Analysis and Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany, [2]Research Department, University of Sofia "St. Kliment Ohridski", Sofia, Bulgaria, [3]Department of Meteorology and Geophysics, University of Sofia "St. Kliment Ohridski", Sofia, Bulgaria

Abstract Data from profiling floats in the Black Sea revealed complex temporal and spatial relationships between physical variables and oxygen, chlorophyll and the backscattering coefficient at 700 nm, as well as some limits in understanding the details of biogeochemistry dynamics. To account for different interdependences between physical and biogeochemical properties, a feedforward backpropagation neural network (NN) was used. This NN learns from data recorded by profiling floats and predicts biogeochemical states using physical measurements only. The performance was very high, particularly for oxygen, but it decreased when the NN was applied to older data because the interrelationships between the physical and biogeochemical properties have changed recently. The biogeochemical states reconstructed by the NN using physical data produced by a coupled physical–biogeochemical operational model were better than the biogeochemical outputs of the same coupled model. Therefore, the use of data from profiling floats, physical properties from numerical models and NNs appears to be a powerful approach for reconstructing the 4D dynamics of the euphotic zone. Basin-wide patterns and temporal variabilities in oxygen, backscattering coefficient and chlorophyll were also analyzed. Of particular interest is the reconstruction of short-lived biogeochemical features, particularly in coastal anticyclone areas, which are difficult to observe with available floats at the basin scale.

Plain Language Summary This study addresses the biogeochemical dynamics of the euphotic layer in the Black Sea. Observations are provided from profiling floats, and the observed biogeochemical parameters include oxygen, the backscattering coefficient at 700 nm and chlorophyll-a. Data analysis showed complex temporal and spatial relationships between physical and biogeochemical variables and some limits in understanding the details of biogeochemical dynamics. A feedforward backpropagation neural network was developed, which can be considered an input–output mapping in which the neurons combine the input data in such a way that the output can be considered a nonlinear combination of input data. When applied to older data, the reconstruction performance decreases, suggesting a change in the dependency of biogeochemical characteristics on physical drivers caused by known climate change. A comparison with simulations from a coupled operational biogeochemical model shows that the neural network outperforms the numerical model. The newly proposed method, combining data from profiling floats, physical properties from numerical models and a backpropagation neural network, allows us to reconstruct the 4D dynamics of the euphotic layer over the period 2013–2020.

1. Introduction

Several important developments in recent decades have greatly advanced our understanding of the biogeochemistry of the ocean and its evolution. The first is attributable to satellites, which continue to provide data on some surface biogeochemical (BGC) properties. The second advancement is the development of sensors for acquiring in situ measurements, particularly by new observational platforms such as the BGC-Argo network (Claustre et al., 2020) and gliders. The third is the advancement in numerical BGC modeling. Unfortunately, each of these advancements suffers from numerous problems. Satellites observe only the ocean surface, whereas the layers beneath could be much more representative of the BGC system than the thin surface layer. BGC-Argo sensors measure properties over large depth ranges; however, sampling locations cannot be determined in advance, and coverage remains insufficient over most of the world's oceans. In addition, the number of sensors is limited.

Gliders can be preprogrammed with travel waypoints; however, they cover relatively small areas of the ocean. Numerical models provide a tremendous amount of data, but they are not optimally realistic. One alternative to these approaches is to develop new analysis methods, for example, machine learning (ML) techniques, which have emerging oceanographic applications. Accordingly, in the present study, we present an example of utilizing ML based on BGC-Argo data from the Black Sea region.

The use of neural networks (NNs) in research on the BGC properties of oceans is not new. Schiller and Doerffer (1999) proposed an algorithm to derive the concentrations of phytoplankton, suspended matter and colored dissolved organic matter over turbid coastal waters using an NN that parameterized the inverse of a radiative transfer model. For this purpose, they implemented data from the Medium Resolution Imaging Spectrometer. Richardson et al. (2002) used ocean color measurements acquired by satellites to map the near-surface phytoplankton distribution and to study the shapes of vertical chlorophyll profiles in the Benguela upwelling system. In the years since these studies, the space–time coverage of surface observations from satellites has continuously improved, and applications aiming to retrieve chlorophyll (Chl-a) concentrations from ocean color measurements have increased in number, including applications based on ML (for extensive reviews, see Sammartino et al., 2020; Sauzède et al., 2015, 2017).

The Black Sea was chosen as the test area because it is small and easily manageable, being an enclosed body of water from which floats do not enter or leave. The complex relationships between the physical and BGC vertical structures in this basin have been extensively investigated (Konovalov & Murray, 2001; Konovalov et al., 2005; Murray & Yakushev, 2006; Murray et al., 1989; Stanev et al., 2004; Tugrul et al., 1992; Yakushev et al., 2007). These and many other studies clearly demonstrated the role of vertical density stratification in the distribution of living organisms and tracers. However, most of these studies focused on the physics of the baroclinic layer, that is, on depths below the euphotic zone.

In the present study, we investigate the photic layer and the upper part of the nitrocline. The photic layer has two oxygen sources: the air–sea oxygen flux (mostly in winter) and the photosynthesis of phytoplankton (mostly in summer). The oxygen sink in the Black Sea is associated with the oxidation of organic matter, which ultimately produces nitrate. A nitrate maximum occurs at the top of the suboxic zone (the zone separating oxic from anoxic waters), from which the concentration of nitrate decreases upward (Konovalov & Murray, 2001; Konovalov et al., 2005). Winter convective mixing triggers a flux of oxygen to the cold intermediate layer from above and the entrainment of nitrate from the upper pycnocline (nitrate pool, nutricline) from below. Of these processes, the latter stimulates plankton growth, the maximum blooms of which occur in late autumn–early winter and spring (Mikaelyan et al., 2017).

While satellite data have been utilized for numerous applications in the oceanography of the Black Sea (Ginsburg et al., 1997; Karabashev et al., 2006; Kopelevich et al., 2004; Kubryakova et al., 2021; Sur et al., 1996), the use of data from profiling floats to study BGC processes is relatively new. Stanev et al. (2013) analyzed oxygen dynamics using data from profiling floats for the first time. This study was soon followed by others using more bio-optical sensors that addressed Chl-a, the particulate backscatter coefficient (bbp) at different wavelengths, and sulfide (Capet et al., 2016; Kubryakov, Mikaelyan, & Stanichny, 2019, 2020; Kubryakova et al., 2018; Stanev et al., 2014, 2017, 2018). However, these studies did not fully reveal the interrelationships between the observed parameters. For instance, the differences in the temporal and spatial variability of oxygen, chlorophyll and the backscattering coefficient are not yet thoroughly understood. The key question is therefore whether the currently available BGC-Argo data are sufficient to reliably describe the basin-wide spatial and temporal variability of these parameters, including their seasonal and interannual variations.

In contrast to some previous analyses of the euphotic layer in the Black Sea, in this study, we aim to (a) analyze the complex relationships between several BGC variables (oxygen, Chl-a and bbp) and (b) investigate the usefulness of ML and its power to derive BGC variables from physical variables. The respective performance of this approach is evaluated against situations excluded from the learning phase. The major novelty is the reconstruction of the basin-wide 4D dynamics of the photic zone. We will demonstrate that the proposed method allows us to derive credible conclusions about basin-wide BGC dynamics of the photic zone. One important difference between the present study, which is focused on spatial distribution of observed properties and their different states in individual years, and some earlier studies is that, in the latter, greater weight was placed on describing the 1-D characteristics of the Black Sea BGC system and focusing mostly on the corresponding seasonal variability.

Moreover, we attempt to employ ML to understand the roles of physical drivers and the conditions under which specific spatial and temporal characteristics appear.

The remainder of this paper is structured as follows: In Section 2, we present the data used and the ML method employed to reconstruct the BGC properties using physical data as an input. Section 3 describes the physical and BGC properties as seen by profiling floats and the performance of the reconstructions. Comparisons of the performance of the ML method with results from the operational coupled physical–BGC model are also given. Finally, we present a 4D reconstruction of the dynamics of the euphotic layer during 2013–2020. In the discussion, we summarize the mechanisms of BGC dynamics in the euphotic zone, their complexity, sampling problems and the justification for using ML as a tool to study basin-wide temporal and spatial variability as well as to provide better operational products. The paper ends with brief conclusions.

2. Data and Methods

2.1. Data

2.1.1. In Situ Data

Data from seven BGC-Argo floats are used in this study. Because profiling floats usually operate in the deep sea, most trajectories (Figure 1a) approximately follow the main cyclonic rim current. BGC-Argo floats are equipped with optodes to measure dissolved oxygen; a multispectral radiometer that measures photosynthetically available radiation (PAR) and downwelling irradiance at 380, 410, and 490 nm; chlorophyll fluorometers; sensors for measuring bbp at 700 nm; and sensors for the measurement of sulfide and nitrate concentrations. In the present study, which addresses the photic layer, we do not use data for sulfide concentrations; they are described by Stanev et al. (2018).

Coccolithophores, which are represented in the Black Sea mainly by *Emiliania huxleyi*, induce strong backscattering throughout the visible spectral range. This characteristic makes them easily observable using optical methods of remote sensing (Kubryakova et al., 2021). In the following, we separately address coccolithophores and phytoplankton/chlorophyll considering that coccolithophores are also planktonic organisms.

More details about the observations, their technical specifications, signal processing, quality control and definitions are given in Appendix 1 and in previous publications (Kubryakov et al., 2020; Rasse et al., 2020, 2021; Roesler et al., 2017; Stanev et al., 2014, 2017, 2018; Xing et al., 2017). The number of floats operating simultaneously is shown in the upper-right inset of Figure 1a. The number of observations provided by each float (Table 1) demonstrates that the volume of recorded BGC data greatly exceeds the volume of previously recorded data. Although BGC floats provide less dense coverage of the basin than floats measuring only seawater physical properties (compare with Figure 1 of Stanev et al., 2019), the former clearly represent unique samples over long time periods covering the entire deep basin.

2.1.2. Data From Numerical Models

Temperature, salinity, oxygen and Chl-a data produced by numerical models have been retrieved from the freely available Black Sea data distributed by the Black Sea—Monitoring Forecasting Centre (BS MFC).

The numerical model used to produce the physical data is based on NEMO general circulation ocean model (version 3.6; Madec et al., 2016), implemented in the Black Sea with horizontal grid resolution of 1/36° zonally and 1/27° meridionally (ca. 3 km) and 31 vertical levels. This resolution resolves the mesoscale eddies (the Rossby radius of deformation in the Black Sea is ~20 km). ECMWF ERA5 atmospheric data at the resolution of 0.25° in space and 1-hr in time are used as atmospheric forcing. The model is online coupled to OceanVar data assimilation scheme to assimilate sea level anomaly along-track observations from the Copernicus Marine Environment Monitoring Service (CMEMS, https://marine.copernicus.eu/). Available in situ vertical profiles of temperature and salinity from both SeaDataNet, which is a pan-European infrastructure to ease the access to marine data measured by the countries bordering the European seas and CMEMS datasets are also assimilated. Further details about the model set up, are given by Ciliberti et al. (2021).

The biogeochemical model data used in this study are produced by means of the BiogeochemicAl Model for hypoxic and Benthic Influenced areas (BAMHBI, Grégoire et al., 2008). This model describes the food web through 24 state variables, several plankton functional types and an explicit representation of the bacterial
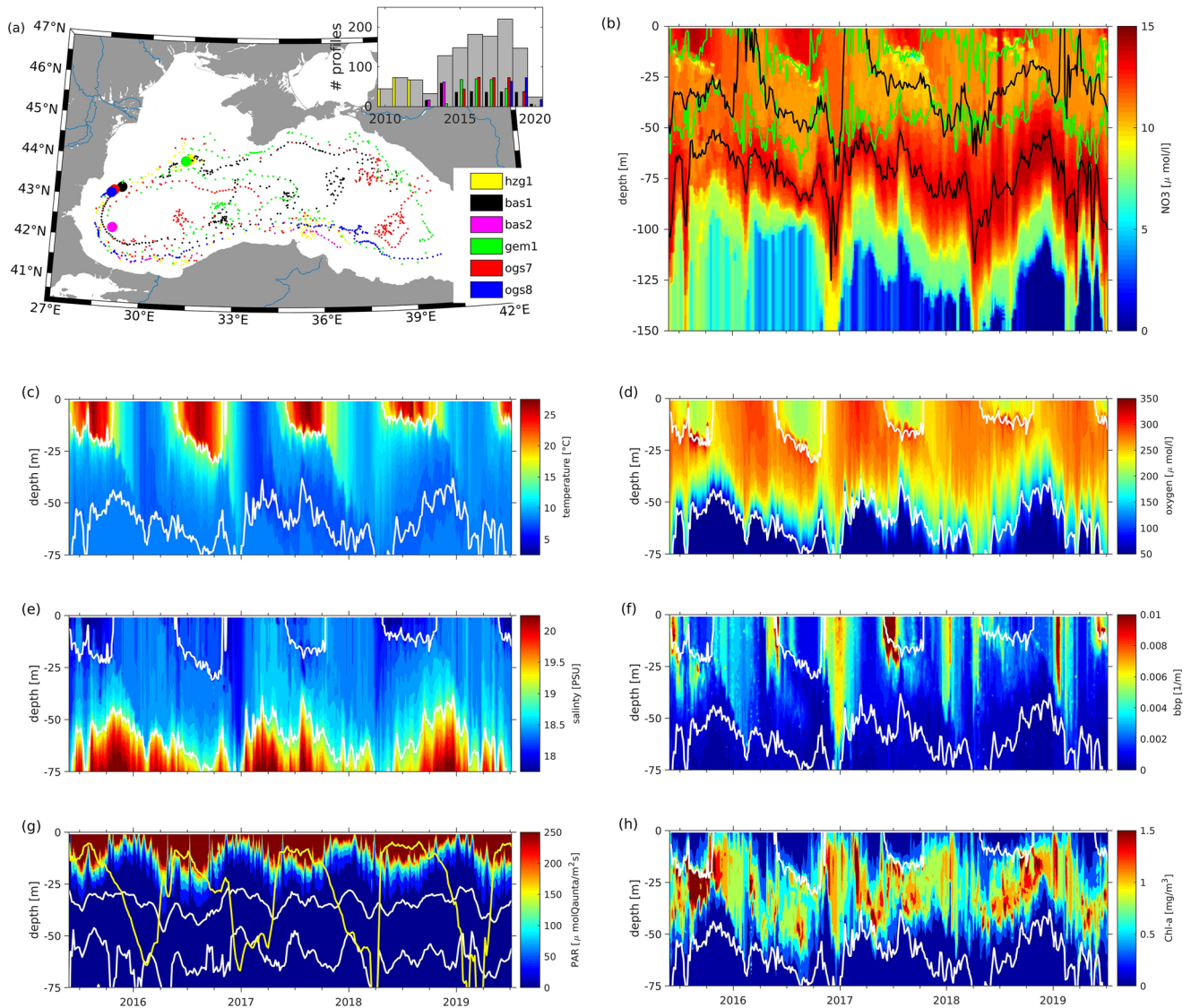
Figure 1. (a) Sampling positions (colored dots) of the BGC-Argo floats operating from June 2009 to March 2020. The deployment positions are shown with larger symbols. The histogram in the inset shows the number of profiles per year sampled by each float as well as the total number of profiles per year (grey). The colors assigned to individual floats are shown in the inset, bottom-right. Time versus depth diagrams (b,…, h) show observed by float ogs7 data. (b) Represents the (detrended) nitrate concentration. The 0.5 mg.m^{-3} Chl-a isolines are superimposed with green lines to show the approximate core of the SCM layer, which is between these two isolines. Black lines represent the positions of $\sigma_t = 13.75$ and $\sigma_t = 15.0$. (c, e, and g). Time versus depth diagrams of temperature (c), salinity (e), and photosynthetically available radiation (PAR) (g) in the euphotic zone. White isolines in the temperature and salinity plots denote $\sigma_t = 12$ and 14.5, respectively. In the PAR panel, white isolines denote 1% and 0.1% of its surface value, respectively; the yellow line denotes the depth of the mixed layer. (d, f, and h) Time versus depth diagrams of dissolved oxygen (d), bbp at 700 nm (f) and Chl-a (h) in the euphotic zone. White isolines denote $\sigma_t = 12$ and 14.5. The marks on the time axes of time versus depth diagrams correspond to 1st January (longer marks), and every third month thereafter (shorter marks).

loop. The model simulates oxygen, nitrogen, silicate and carbon cycling. In the frame of CMEMS, BAMHBI is coupled with NEMO.

2.2. Feedforward Backpropagation Neural Network

In this study, we employ a feedforward backpropagation NN (Schiller, 2000) to derive BGC variables from physical variables. This method can be understood as a kind of input–output mapping in which the neurons combine the input data in such a way that the output can be considered a nonlinear combination of input data. This ensures that, unlike some statistical methods, for example, EOF analysis, signal processing can "capture"

Table 1
Overview of the Data Used Herein Originating From Different Floats

| WMO number	7900465	7900466	7900591	7900592	6900807	6901866	6903240
Nickname	awi1	hzg1	bas1	bas2	gem1	ogs7	ogs8
Start date	June-2009	June-2009	Dec-2013	Dec-2013	Dec-2014	May-2015	Mar-2018
End date	Mar-2012	Nov-2012	Feb-2020	Oct-2014	Sep-2018	Jul 2019	Still operating
Observation interval	Weekly	Weekly	10 days	44 profiles daily, then 10 days	5–6 days	5 days	5 days
# **profiles** sampled	**134**	**185**	**264**	**79**	**259**	**302**	**155**
# T, S, p: **euphotic**	**2,058**	**6,713**	**37,196**	**11,218**	**194,808**	**150,729**	**21,562**
(first 10 m)	(401)	(1,139)	(5,764)	(1,738)	(95,577)	(31,905)	(3,082)
Total	16,783	46,657	522,573	138,742	494,398	486,427	119,614
# Oxygen **euphotic**	**1,456**	**6,713**	**36,986**	**10,091**	**19,212**	**37,509**	**6,853**
(First 10 m)	(292)	(1,139)	(14,616)	(4,065)	(10,986)	(11,897)	(2,250)
Total	13,014	46,657	261,904	76,682	109,586	134,705	67,913
# Chl-a, bbp **euphotic**	-	-	**22,776**	**2,276**	**7,233**	**7,647**	**299**
(First 10 m)	-	-	(3,264)	(747)	(1,892)	(541)	(23)
Total	-	-	129,999	33,467	153,649	184,962	75,600
# Nitrate **euphotic**	-	-	-	-	-	**7,390**	-
(First 10 m)	-	-	-	-	-	(1,327)	-
Total	-	-	-	-	-	15,145	-

Note. Acronyms correspond to the name of the operating institute (awi-Alfred Wegener Institute, hzg-Helmholtz Centrum Geesthacht, bas-Bulgarian Academy of Science, gem-GeoEcoMar, ogs-Istituto Nazionale di Oceanografia e di Geofisica Sperimentale). The last four lines present the numbers of samples in the euphotic zone: from the surface to 70 m (bold number), in the upper 10 m (in brackets), and the total number of samples (the number at the bottom of each line).

the nonlinear interactions between processes with different time scales, that is, the method is able to reproduce nonlinear dynamics.

The decision to use a feedforward backpropagation NN is justified by the limited number of profiles from BGC-Argo floats (only a few hundred). On the other hand, the number of measurements within a given profile is very large; however, the vertical resolution varies among different profiles/floats, leading to nonuniform vertical sampling (see Table 1 and Appendix 1). The aforementioned feedforward backpropagation NN is well suited for this task, as these NNs can work with pointwise (0-D) measurements rather than processing whole profiles at once. In contrast, the usage of more advanced ML techniques, such as deep learning (a broader range of ML methods), which can process 1-D profiles (Kiranyaz et al., 2021), would necessitate data augmentation.

NNs are organized into layers (Bishop, 1995): one input layer, one output layer and one or more hidden layers in between (Figure 2). Each layer consists of neurons, and the numbers of neurons in the input and output layers are predefined by the training data, whereas the architecture of the hidden layer(s) depends upon the complexity of the problem. Each neuron in a layer is linked to each neuron in a neighboring layer by weighting. NNs work sequentially: each element of the input vector serves as the input for one of the neurons of the input layer. The output of the first hidden layer is computed by summing the weighted inputs, shifting the sum by a bias, and applying a nonlinear function (a sigmoid activation function in this study). This procedure repeats itself until the output layer is reached, where the outcome of each neuron gives one element of the output vector. The weights and biases are free of the approximation; that is, they are fixed during the training phase of the NN. The mean squared relative error per neuron is iteratively minimized during

$$y_j = f\left(-b_j + \sum_i w_{ij}x_i\right)$$

$$f(x) = \frac{1}{1 + \exp(-x)}$$
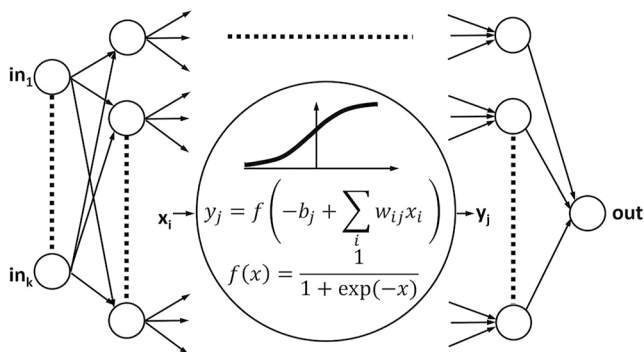
Figure 2. Neural network (NN) working scheme: input information (ARGO physical data, some auxiliary data) is sequentially passed through the network using a nonlinear activation function (the sigmoid activation function is used herein) until the output layer is reached. The NN output is oxygen, bbp or Chl-a. The free parameters of the NN neurons (weights w_{ij} and biases $b_j w_{ij}$) are adjusted during the NN training phase with a gradient descent scheme.

the training by backpropagation through the NN and adjusting the biases and weights according to a gradient descent scheme.

3. Results

### 3.1. Physical and BGC Properties as Seen by Profiling Floats

The data collected by float ogs7 are shown in Figure 1 as a representative example of time versus depth diagrams and because this is the only float measuring nitrates. On average, PAR drops to 1% of its surface value at a depth of ∼35–40 m (Figure 1g) and is between 1% and 0.1% of its surface value at 40–70 m. In Figure 1, we display only the data within the first 70 m, which includes the euphotic zone, defined as the layer where PAR is more than 1% of its surface. Only for nitrate are deeper layers shown in Figure 1b to better display the zone of nitrate maximum and nitrocline. The yellow line in Figure 1g shows the mixed layer depth (MLD), computed using the method suggested by Kara et al. (2009). These authors define MLD as the depth where the density has changed by an amount of dr from the density at a reference depth of 3 m. Following Kubryakov, Belokopytov, et al. (2019) dr has been taken as equal to 0.07 kg/m^3.

The temperature of the shallow mixed layer exhibits rather regular seasonal variability; however, this variability differs strongly in individual years (Figure 1c) as well as in the different locations visited by the floats. We remind the reader that the characteristic variations and individual events in Figure 1 to which we refer can be identified by the time of their occurrence. This time can be seen on the time-axes. The same convention applies to all time versus depth diagrams presented in the following. The year with the strongest cooling in the last decade was 2017, which is well illustrated by the outcropping of the density surface $\sigma_t = 14.2$ (upper white lines) and the downward propagation of the associated cold water signal. The temporal variability in salinity (Figure 1e) is much less regular than that in temperature. During cooling events, vertical mixing contributes to an increase in surface salinity. The extremely strong event in 2017 (Figure 1c) gives one example of such a scenario.

The density varied over a relatively large range. Above the base of the upper mixed layer, the large range of density is explained by the large range of temperature, whereas below this depth, the large range of salinity explains the range of density variations (see also Stanev & Chtirkova, 2021). This characteristic is very important for reconstructing the BGC states from physical data because temperature variability is strongly related to the illumination conditions, and the variability in salinity at the bottom of the photic layer reflects the dynamics of the pycnocline. Thus, the dynamics of the pycnocline govern the position of the nitrocline (Figure 1b); therefore, it is considered a strong predictor of the properties of the phytoplankton community. In conclusion, the temperature-stratified upper layer and salinity-stratified depths below 50 m provide the ecological niche responsible for the major variability of the BGC system in the photic layer.

The nitrate maximum zone nicely follows the position of $\sigma_t = 15.0$. The correlation between the concentration of nitrate and density (full line in Figure 3a) clearly shows the role that baroclinic dynamics play in the temporal and spatial variability of nitrate below 35 m. The anti-correlation reaches a maximum of 0.8 in the pycnocline (at ∼100 m), illustrating that the upwelling of deeper (nitrate-poor and denser) water reduces the nitrate concentration at this depth. Above the nitrate maximum zone, upwelling introduces nitrate-rich water into the upper layer, and the correlation with density increases up to 0.5 at a depth of 55 m. In the shallow seasonal pycnocline, the correlation again becomes negative (∼-0.9) at the surface. The subsurface chlorophyll maximum (SCM) layer shows rather complex temporal and spatial variability (Figure 1h). The two green isolines in Figure 1b, showing a Chl-a concentration of 0.5 mg.m^{-3}, approximately represent the relationship between the nitrate and Chl-a profiles.

The density isolines in Figure 1d show a clear correlation between the concentration of oxygen and the position of the pycnocline. Obviously, while temperature is the driver for the oxygen variability in the upper layer (higher saturation at lower temperatures in winter), the dynamics of the pycnocline shape the distribution of oxygen at the bottom of the photic zone. A subsurface oxygen maximum occurs in summer, which is fueled by photosynthesis in the SCM layer. This statement is supported by the analysis of saturation ratio So_2 in Figure 4, expressed as the ratio of measured oxygen concentration to the oxygen concentration of water at equilibrium with the atmosphere based on potential temperature and salinity. Differences in the oxygen patterns in Figures 1d and 4 suggest strong biological activity in the warm part of the year (primary production liberates oxygen). Subsurface oxygen maximum at approximately $\sigma_t = 12$ is much more pronounced in the saturation ratio than in the concentration of
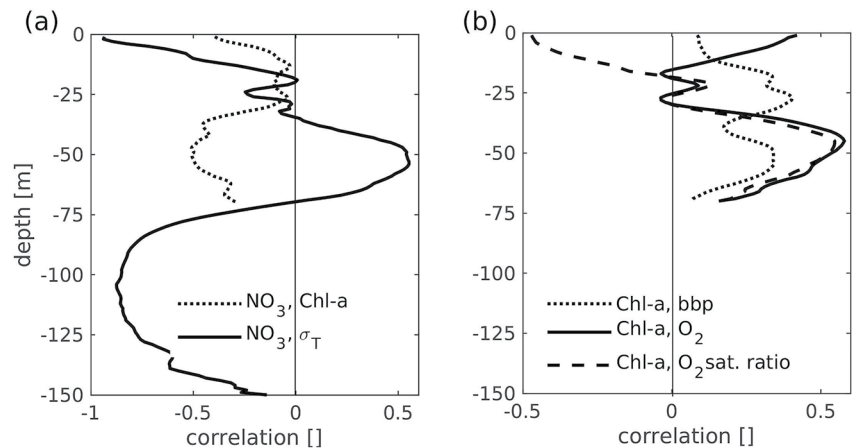
Figure 3. (a) Correlations between nitrate and density (full line) and between nitrate and Chl-a (dotted line). (b) Correlations between Chl-a and oxygen (full line) and between Chl-a and bbp (dotted line). Long-dashed line shows the correlation between Chl-a and saturation ratio. Note that the correlations with Chl-a are shown in only the euphotic zone because the Chl-a values are too small in the deeper layers.

oxygen. Furthermore, in the shallow surface layer studied here concentrations are at approximately atmospheric saturation (saturation ratio is close to unity). Unlike saturation ratio, oxygen concentrations (Figure 1d) reveal nicely the winter maximum associated with surface cooling and winter convection.

Analysis of bbp (Figure 1f) reveals two maxima every year: during winter into the beginning of spring (December to March) and during summer (May to August). Both of these maxima originate at the surface and are in phase throughout the entire basin. Unlike the coccolithophore blooms, the Chl-a maximum (Figure 1h) does not disappear throughout the year; rather, the position of its core changes considerably (~30 m). The upper boundary of the Chl-a layer approaches the sea surface in winter. Afterward, it continuously descends until approximately mid-year and ascends until the next convective event. The bottom of the SCM layer follows the fluctuations of the density surfaces (white isolines in Figure 1h) but not as closely as the constant oxygen lines. In response to the two weak convective events in 2015 and 2016, the bbp and Chl-a signals fluctuated rather strongly. Their magnitudes were comparable to those created by the extreme event in 2017, suggesting a nonlinear response. The overall low correlation among oxygen, bbp, and Chl-a (Figure 3b) reflects how difficult it is to draw simple conclusions about the dominant relationships between these properties. Within the depth range studied herein, a depth of 50 m is particularly important, as the maximum negative correlation between nitrate and Chl-a is "paired" with the maximum positive correlation between nitrate and density (Figure 3a). Slightly above this depth, at ~45 m, the correlation between oxygen and Chl-a appears to reach a significant value of ~0.6. The opposite sign of correlation between the saturation ratio and Chl-a in the upper 25 m (compare the two dashed lines in Figure 3b) gives another evidence about the different inter-relationships between two oxygen characteristics and Chl-a. In the entire depth interval studied herein (down to 70 m), the correlation between bbp and Chl-a remains below ~0.4 (Figure 3b), which largely differs from the known oceanic cases.
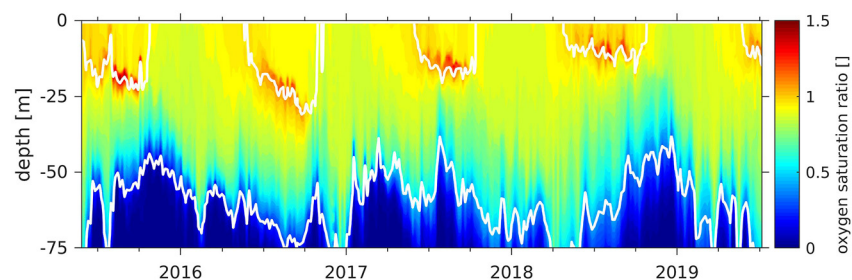
Figure 4. Time versus depth diagram of oxygen saturation ratio observed by float ogs7. White isolines denote $\sigma_t = 12$ and 14.5, respectively.
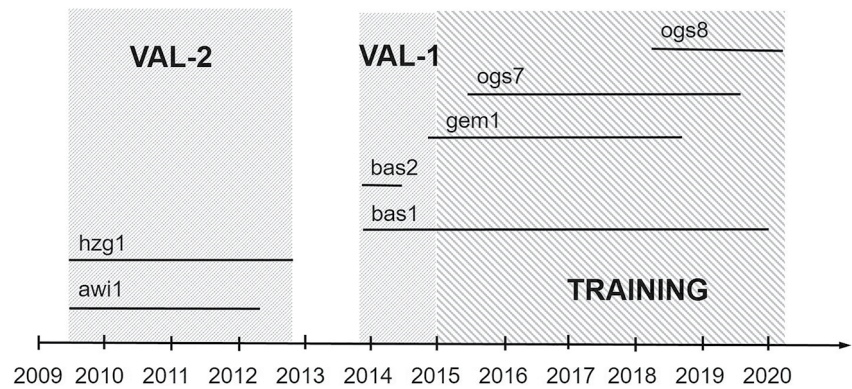
Figure 5. Time flow of data used for training (TRAINING) and validation (VAL-1 and VAL-2). Different shading is used to more clearly show the three individual periods.

3.2. The Network Architecture and Training

Assuming that functional relationships exist between the physical (input) and BGC (output) variables at a given time, position and water depth, an NN with at least one hidden layer can approximate any continuous function (universal approximation theorem), as shown by Hornik (1991). Experimentation with different NN structures was performed to find the most appropriate network architecture. In the first series of experiments, we studied the feasibility of deriving Argo-measured BGC variables (NN output: Chl-a, bbp and O_2) from physical variables (NN input: T and S) and investigated the sensitivity of the reconstructed BGC variables to different combinations of input/output variables and to different network architectures. The important findings of these preparatory works are as follows:

- An NN with two hidden layers is superior to an NN with only one hidden layer. To avoid overtraining, an NN with more than two layers was not evaluated. In this context, the volume of training data is limited, whereas the number of free parameters during NN training increases significantly with an increasing number of hidden layers.
- Spatial and temporal determinations have different impacts on the quality of reconstruction. While the sampling position (latitude/longitude/depth) is a key variable for deriving bbp, day-of-the-year information is crucial for deriving Chl-a. The latter is explained by considering the clear seasonal cyclicity of Chl-a (right panels in Figure 1h).
- The NN performance is significantly enhanced when information on the vertical gradients of physical variables is added and further improved when the vertical gradient is defined by three points rather than by two.
- Employing oxygen as input to the NN does not improve the quality of the derived Chl-a and bbp. Because oxygen cannot be considered a physical parameter, we used it as an NN output variable.
- There is no significant difference in the performance of the NN when we derive the individual BGC parameters simultaneously compared to when separate NNs are built for each individual BGC parameter.
- Including σ_t as an input variable only slightly improves the NN performance.

The above conclusions led us to choose the following information for the NN input layer: (a) locations and times of the Argo profiles specified by the latitude, longitude, and day-of-the-year; (b) vertical position and physical variables at this position (depth z, temperature $T(z)$, salinity $S(z)$, and density defined as $\sigma_t = \rho(S,T)-1,000$ kg m^{-3}); and (c) information about the vertical gradients of the physical variables within this Argo profile given by the temperature, salinity and density at three depths (2, 20, and 50 m). In total, the NN has 17 input variables and 1 output variable: O_2, Chl or bbp. Thus, in total, the NN consists of 4 layers with 17, 17, 7, and 1 neurons in them. A network, in which So_2 was used instead of O_2 gave almost identical results, which is explained by the fact that the network learns perfectly the dependency of oxygen concentration of water at equilibrium with the atmosphere on T and S.

The NN was trained using all data from all BGC-Argo floats since 1 January 2015 (TRAINING-period in Figure 5, see also Table 1). Data prior to 1 January 2015 (VAL-1 period) were kept as an independent validation data set. During the second validation period (VAL-2 in Figure 5) validation of only oxygen data was performed because

Table 2
RMSREs of the NN Reconstruction of Oxygen, bbp, and Chl-a After Training

Variable	RMSRE (train)	RMSRE(train)/RMSRE(test)
Oxygen	2.70%	0.91
bbp	2.69%	1.14
Chl-a	1.55%	1.02

Note. The ratios of the RMSREs between the training and testing data are also given.

no other BGC data were available. The training data set consists of 120,401 samples for Chl-a and bbp and 89,418 samples for oxygen. The validation data set is approximately 6 times smaller; we admit that the total number of data points did not allow the validation data set to be larger. During the NN training phase, a fraction of the training data (∼5%) was used as an independent testing data set (i.e., they were not used to determine the free parameters of the NN) to avoid overtraining the NN. The training of the NN was supervised by keeping track of the error as a function of a training epoch. The error minimized during training is the root mean square relative error (RMSRE) since all variables are scaled to the range [0, 1]. The RMSREs of the training points and the ratio of the RMSREs between the training/testing points after training shown in Table 2 demonstrate that the training quality was good. Please note that the numbers in Table 2 yield only rough information about the training quality. These data are not fully suitable for comparing the reconstruction capabilities due to the very different ranges of the original data sets.

3.3. Validation of Results

We first analyze the goodness of fit of the derived BGC properties by comparing the observations and emulations for the training period. During this period, the NN computes an optimal state by minimizing the differences between the predicted (derived) values and the values observed by individual floats. The following index of agreement is used to estimate how well the reconstructed data match the observations (Willmot, 1981):

$$D(P, Q) = 1 - \sum_{i=1}^{n} (P_i - O_i)^2 / \sum_{i=1}^{n} \left(|P_i - \overline{O}| + |O_i - \overline{O}| \right)^2, \tag{1}$$

where P_i and O_i are the BGC variables from the two data sets (predicted and observed, respectively) at an individual depth, n is the number of observations at this depth (the index is i), and the overbar indicates the mean for each depth. D measures the degree to which P matches O, where 1 indicates perfect agreement and 0 indicates no agreement at all. For the upper 70 m layer, the average indices are 0.97, 0.91, and 0.83 for oxygen, bbp and Chl-a, respectively.
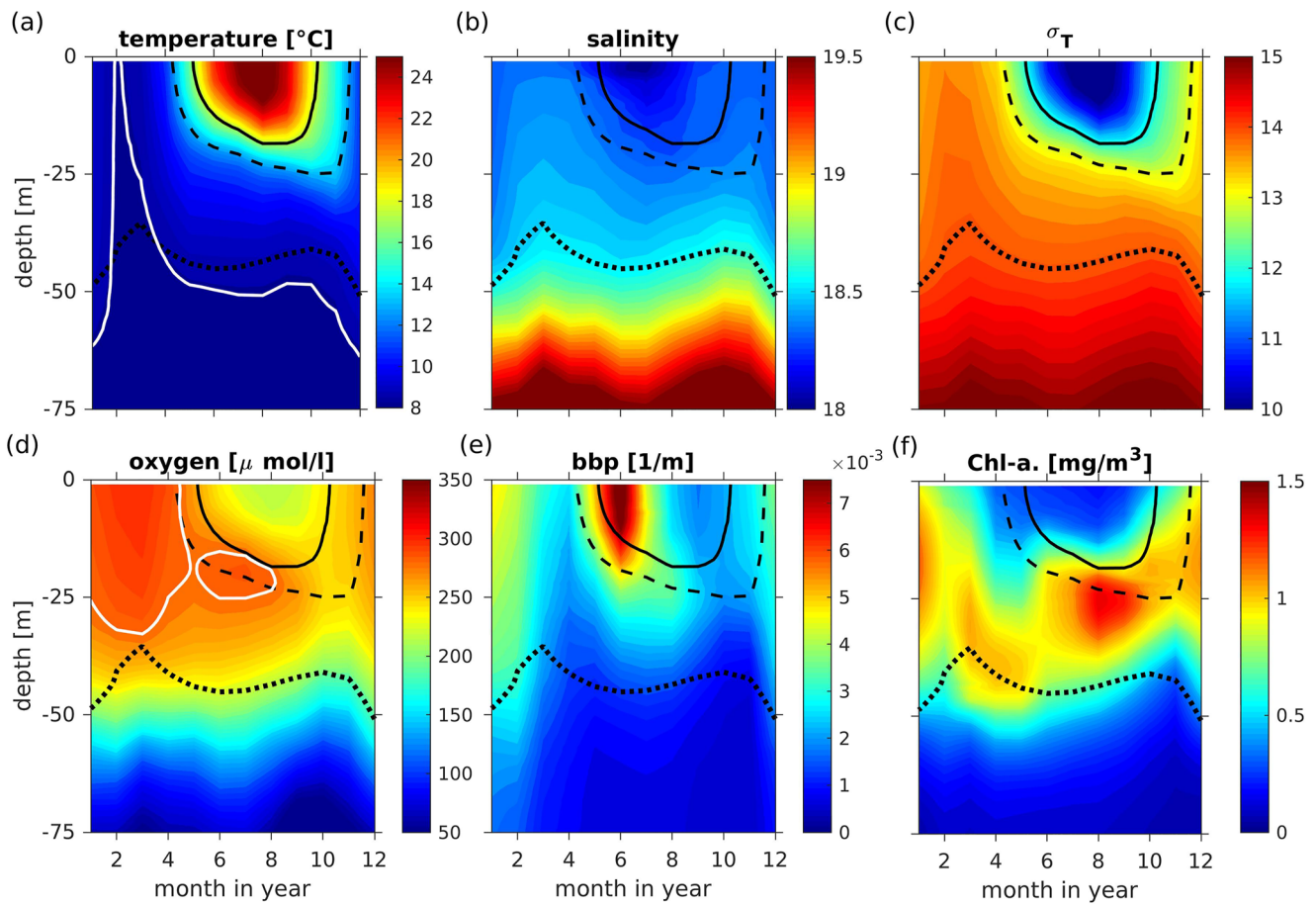
The time versus depth diagrams in Figure 6 present the physical (observed) and BGC (reconstructed) data monthly averaged over the training period. Oxygen ventilation of the upper layers associated with winter cooling penetrates down to ∼40 m, and the propagation of this signal into the deeper layers is opposed by the vertical diffusion and mixing of sulfide-rich water from the anoxic layers below.

In the depth range studied here, bbp monotonically decreases with increasing depth. The time versus depth diagram of bbp is approximately the opposite of that of density, particularly below 50 m, where the upward propagation of denser waters from below parallels the upward propagation of low-bbp waters. The two bbp maxima in November to-February and April to August penetrate to very different depths.

The SCM layer has a thickness of ∼30 m, reaches the sea surface in winter and submerges below the upper mixed layer in summer. Notably, the seasonal evolution of the SCM (Figure 6f) does not exhibit the "boot heel" shape in winter reported in earlier modeling works (e.g., Figure 5 of Oguz et al., 1996; see also Kubryakova et al., 2018). By "boot heel" shape we mean that the time versus depth plots of phytoplankton concentration show an abrupt overshooting in the downward displacement of isolines in winter, which is followed by a slow upward displacement and then an even slower shoe sole-like displacement.

The depths where the SCM layer is usually observed are overshot only during the period of extremely strong winter cooling in 2017, when the vertical mixing registered by floats gem1 and ogs7 reached depths exceeding 50 m (Figure 1). A comparison of the indices of agreement between the observed and derived monthly data (0.99, 0.93, and 0.98 for oxygen, bbp and Chl-a, respectively) demonstrates that the NN learns to almost perfectly reproduce the seasonal variability of the observations.

To illustrate the performance of the NN for most of the training period, Figure 7 presents the observed (left panels) and reconstructed (middle panels) data for float bas1. Obviously, the oxygen saturation ratio is reproduced exceptionally well throughout the water column and during all seasons and years. The timing and vertical
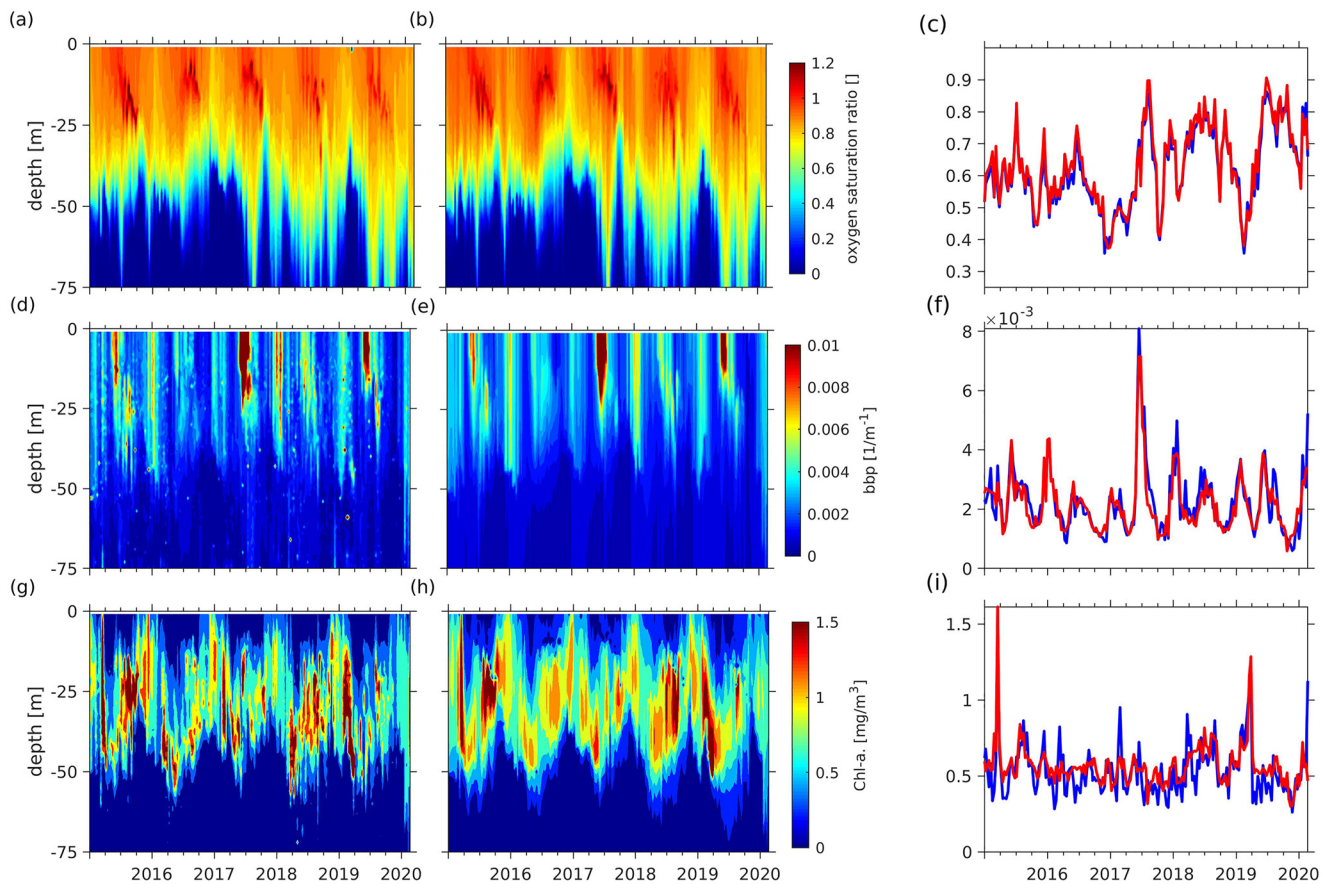
Figure 6. Time versus depth diagrams of the observed monthly mean physical (top) and Neural network (NN)-derived BGC (bottom) variables during the training period. The averaging includes all data from the floats operating during the respective months. The white isoline (temperature 8.5°C) in (a) denotes the position of the upper boundary of the cold intermediate layer. The white isoline (280 µmol/l) is superimposed in (d). Black full, dashed and dotted lines are density surfaces 12, 13, and 14, respectively.

extent of the bbp maxima are also very well reproduced. The complex distribution of the Chl-a content throughout the water column was the most challenging, but this parameter was also well replicated in the NN reconstruction (compare Figure 7g with Figure 7h). The positions and maximum values of the SCM are met at most times.

The graphs on the right-hand side of Figure 7 show the properties integrated over the upper 70 m for bas1 as observed and reconstructed by the NN. Overall, the NN tends to increase the values when other simultaneously operating floats observe values higher than those detected by bas1, and vice versa. The goodness of NN reconstruction for bas1 estimated by the indices of agreement for oxygen, bbp and Chl-a is 0.95, 0.80, and 0.81, respectively. Obviously, the reconstruction of interannual variability is more challenging than the reconstruction of seasonal cyclicity. One explanation for this is that the learning period is not long enough to account for the interannual variability. Another important message conveyed by Figure 7 is that while the reconstructed oxygen saturation ratio and bbp (vertically averaged over the upper 70 m) effectively match the observed values (Figures 7c and 7f), the vertically averaged Chl-a values follow the interannual signal less well (Figure 7i). This finding demonstrates that the variability of the vertically averaged Chl-a is not well coherent with temperature and salinity. Indeed, the occurrence time of the SCM and its depth are consistent; therefore, the observations and derivations still exhibit good agreement ($D = 0.81$).

As explained when presenting the NN, we retained the data observed before 1 January 2015 as a validation data set (VAL-1 period in Figure 5). The reconstructed BGC data agree qualitatively with the observations. The means for the upper 70 m layer indices of agreement of oxygen, bbp and Chl-a are 0.95, 0.58, and 0.49, respectively, for BGC-Argo float bas1 and 0.86, 0.56, and 0.42 for BGC-Argo float bas2. Obviously, the NN learned enough to

Figure 7. Observed and reconstructed data from float bas1 during the training period for (a, b, and c) oxygen saturation ratio (d, e, and f) bbp, and (g, h, and i) Chl-a. The panels on the left show time versus depth diagrams of the observed properties, the panels in the middle show the diagrams for the reconstructed properties, and the panels on the right show the diagrams for the properties averaged over the upper 70 m from the observed data (blue) and reconstructions (red).

reconstruct the oxygen profile. Regarding bbp, the timing of the summer 2014 bloom is captured, but the temporal and spatial evolution of Chl-a is only roughly derived.

Profiling floats awi1 and hzg1 provide additional validation data, as they cover an interval that is approximately twice as long (VAL-2 period in Figure 5). However, these floats measured only oxygen data. The validation is presented in Figure 8 for the hzg1 float; the index of agreement is 0.76. The lower skill corresponding to these earlier data would indicate either problems with the sensors or changes in the controlling influence of physical drivers on the BGC characteristics during recent years (see Section 4). The second conjecture is supported by the trend toward lower oxygen concentrations and oxygen saturation ratios in the reconstructed properties, that is, the NN is biased by the deoxygenation trend of recent years.

3.4. Machine Learning Versus Numerical Simulations Using Coupled Modeling

Stanev and Chtirkova's (2021) exercise of inferring seasonal mean physical fields on a relatively coarse grid of 0.5° showed that even for physical properties (many more floats), the data volume needs to be further increased to ensure that the interpretations are credible. However, profiling floats are not the only source of physical data in the interior of the Black Sea. Thus, a question arises as to whether we can use the available data from numerical (physical) models as inputs for the developed NN to reconstruct the BGC states corresponding to the model physics. In this context, we can ask the following question: how does the reconstruction skill using the approach developed in the present study compare with the skill of a coupled physical-BGC model?

The solution to the problem formulated above is facilitated by the availability of data from the Black Sea operational model of the CMEMS. This model assimilates altimeter data and physical data from profiling floats;
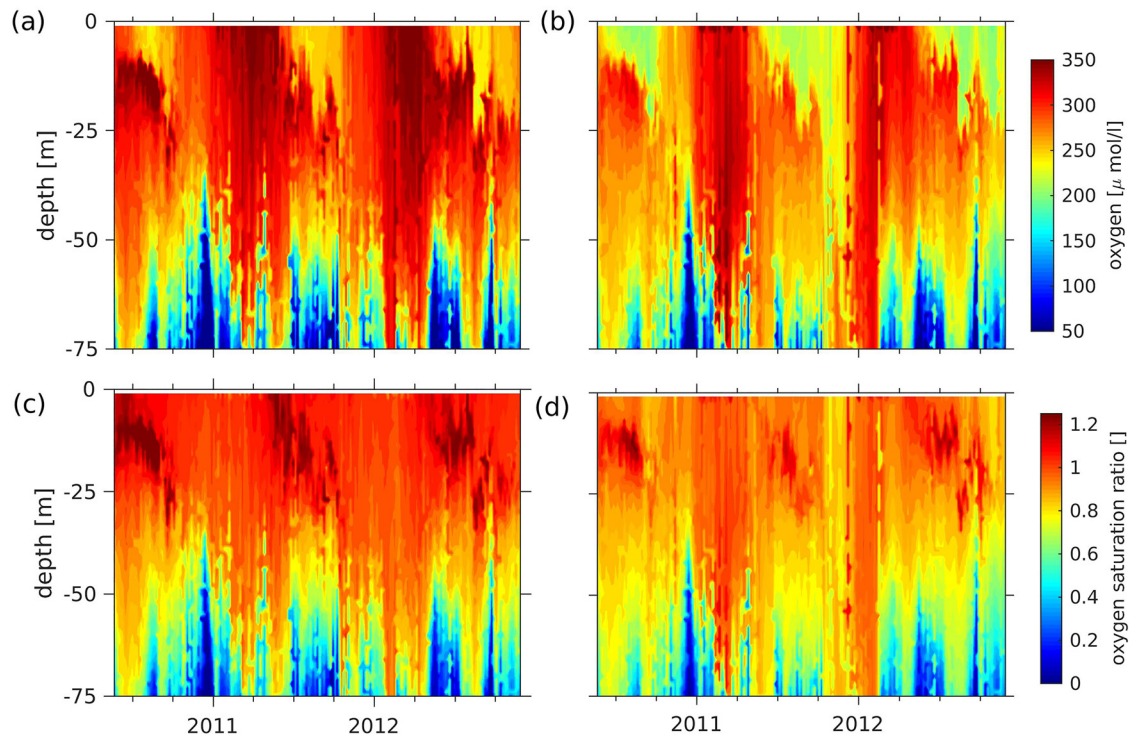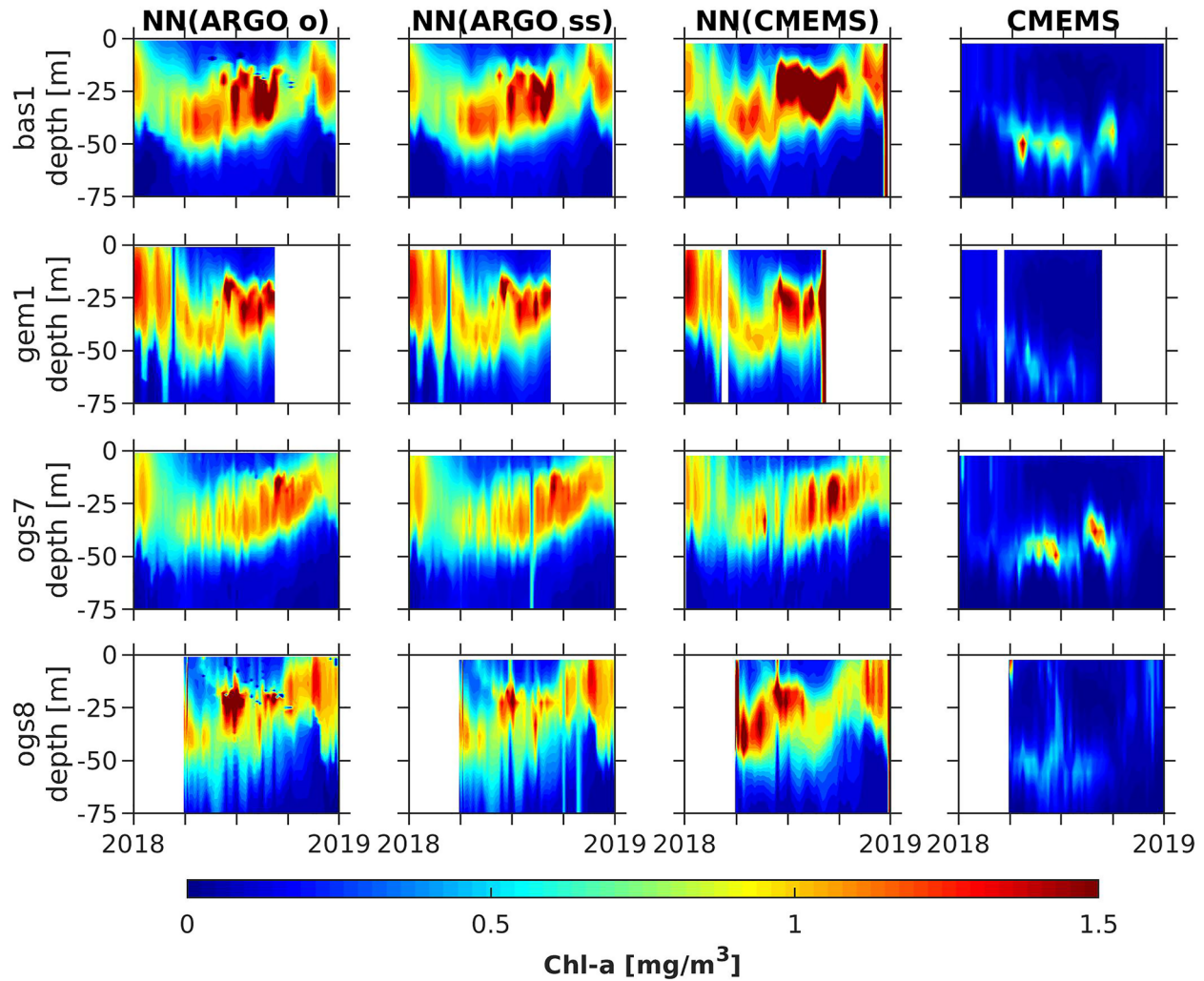
Figure 8. Time versus depth diagrams of observed (a, c) and Neural network (NN) reconstructed (b, d) oxygen (top) and oxygen saturation ratio (bottom) in the upper 70 m layer from the hzg1 data.

therefore, it is expected that the 3-D physical fields do not deviate much from the observations. As a test year, we chose 2018, which is within the time interval over which we trained the NN. During 2018, four floats operated simultaneously; gem1 and ogs8 were active only during part of the year. One problem is that the vertical resolution of model data is much coarser than the vertical sampling resolution of profiling floats. Therefore, the first experiment that we performed aimed to estimate the impact of the vertical resolution on the reconstructed BGC fields. Two reconstructions were performed using the physical data from the profiling floats: one, as described above, using the full BGC-Argo physical data set and a second using subsampled data at the times (positions) when (where) the profiles were taken but at the depths of the model levels. The panels on the left-hand side in Figures S1 and S2 in the Supporting Information S1 and Figure 9 show the NN reconstructions for oxygen, bbp and Chl-a, respectively, using the original data (Argo_o). The panels in the second column show the NN reconstructions using the subsampled CMEMS model-depth data (Argo_ss). The four lines in this figure show the corresponding patterns from the four floats operating in 2018. The results of the comparison of the NN reconstructions in the two individual experiments against the observed data are summarized in Table 3 (the first and second columns). The index of agreement remains nearly the same in the two reconstructions. In summary, a coarser vertical sampling resolution does not substantially affect the goodness of the reconstruction.

The above result justified the next experiment, in which we used physical data from the model instead of the observed data. The reconstructions (in the third columns of Figures S1 and S2 in Supporting Information S1 and Figure 9) show good agreement with the reconstructions based on the subsampled Argo observations (patterns in the second column). The respective indices of agreement and their comparison against the indices when the Argo data on the model levels were used (Table 3) reflect the sensitivity of the responses of the BGC system to different physical forcings. The largest deterioration in the reconstruction quality was observed by oxygen, which was almost perfectly reconstructed from the BGC-Argo observations.

When critically estimating the skill of the developed model to provide reasonable reconstructions of the BGC system, it is instructive to analyze the agreement between the CMEMS BGC simulations and Argo observations at the same positions and times (last columns in Figure S1 in Supporting Information S1 and Figure 9). No

Figure 9. Performance of the neural network (NN) using physical data from BGC-Argo or CMEMS as the input when predicting the Chl-a evolution along the float trajectories presented as time versus depth diagrams for 2018. The panels in the individual columns show: (1) NN reconstruction using the BGC-Argo physical data at their original sampling depths, (2) NN reconstruction using the BGC-Argo physical data at the depths of the CMEMS model levels, (3) NN reconstruction using the CMEMS physical data, and (4) data from the coupled physical-BGC model of CMEMS. Each of the four lines shows the results for one of the operating floats. Argo_o and Argo_ss denote the data at the original sampling depths and the data subsampled at the model levels, respectively.

comparison is shown here for bbp because the CMEMS model does not provide such data. Obviously, for all floats and parameters, the NN reconstructions outperform the results of the model simulations (see also Table 3).

3.5. 4D Reconstruction of BGC Dynamics

The results presented above motivated us to compute the 4D BGC fields using 4D temperature and salinity data as inputs from the CMEMS model. To demonstrate the level of credibility of the reconstructions, we first show in Figure 10 the time versus depth diagrams of oxygen and Chl-a as modeled by CMEMS-BGC (on the left-hand side) and as emulated by the NN when applied to CMEMS physics (on the right-hand side) during 2015–2020. Data were sampled from the 4D fields along the paths of float bas1. Similar plots for floats gem1, ogs7 and ogs8 are shown in Figures S3, S4, and S5, respectively. A comparison of Figures 10 and 7 clearly shows that the reconstructions when using the NN (panels on the right-hand side of Figure 10) agree much better with the observations (panels on the left-hand side of Figure 7), as well as with the reconstructions based on physics from profiling floats (panels in the middle of Figure 7), than the numerical simulations with the coupled BGC model (panels on the left-hand side of Figure 10).

Table 3

Mean Index of Agreement During 2018 Averaged Over the Upper 70 m Between the Observed BGC-Argo Float Data and the NN Reconstructions Derived Using Physical Data With a Different Origin or a Different Vertical Resolution as the Input

|	NN(ARGO_o)	NN(ARGO_ss)	NN(CMEMS)	CMEMS (bio)
Oxygen				
bas1	0.92	0.92	0.64	0.42
gem1	0.92	0.92	0.66	0.51
ogs7	0.93	0.94	0.83	0.56
ogs8	0.90	0.90	0.62	0.40
bbp-				
bas1	0.73	0.73	0.54	-
gem1	0.92	0.92	0.81	-
ogs7	0.84	0.89	0.54	-
ogs8	0.77	0.77	0.51	-
Chl-a				
bas1	0.76	0.76	0.48	0.29
gem1	0.82	0.79	0.48	0.47
ogs7	0.71	0.79	0.53	0.25
ogs8	0.71	0.72	0.42	0.41

Note. The last column shows the index of agreement for the CMEMS BGC model data. Argo_o and Argo_ss denote the observed (o) and subsampled (ss) Argo data.

As illustrations of 4D reconstructions, we present in the panels on the left-hand side of Figure 11 the mean oxygen at 50 m and bbp and Chl-a at 10 m for the period 2015–2020. The oxygen distribution illustrates the basic dynamics in the Black Sea, which is the cyclonic gyre encompassing the basin (Figure 11a). In the central part of the Black Sea, oxygen-poor water upwells, which explains the low oxygen values in the central areas of the sea.

The annual mean distributions of bbp and Chl-a show very different patterns. Large values of Chl-a along the western coast (Figure 11e) are known from the early days of satellite oceanography (Barale & Murray, 1995). These values are explained by the inflow from the large rivers bringing more nutrients into the system. Our estimations show that, overall, Chl-a concentrations along all coasts are higher than those in the interior of the basin. Here, we note that the results beyond the bathymetric line of 500 m can be considered ML extrapolation because these areas have not been visited by the floats. In contrast to Chl-a, the bbp patterns show maxima along the eastern coasts and in the shallow northernmost parts of the sea (Figure 11c).

The snapshots on 1 January 2018 (on the right-hand side of Figure 11) prove that the proposed method is capable of replicating eddy-like patterns, which are well pronounced along the continental slope, as well as in the basin interior. The strong gradients of BGC properties associated with eddies explain the rapid changes in the observed by the floats characteristics (Figure 1), that is floats cross multiple BGC fronts (see discussion in Section 4.1). Furthermore, these patterns support the results of the coupled physical and BGC numerical model for the Black Sea region of Stanev et al. (2014), which evidenced the role of coastal eddies in the BGC dynamics of this basin.

The 4D reconstruction makes it possible to analyze the temporal changes in basin-wide BGC properties (Figure 12). While oxygen at the sea surface
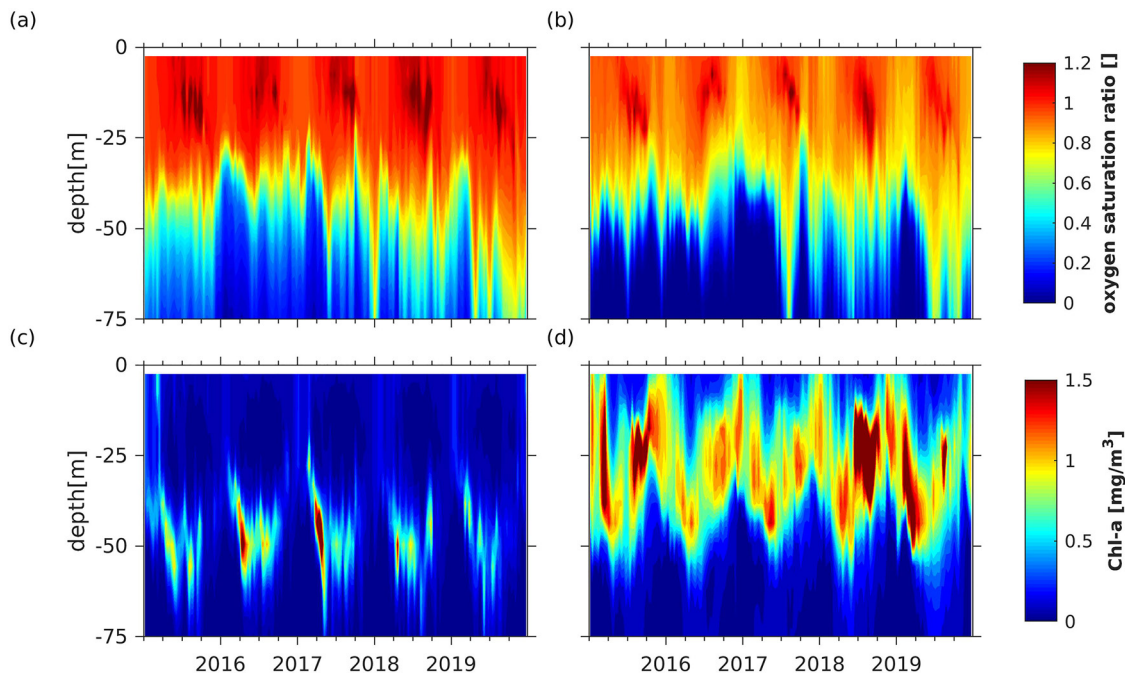

Figure 10. Time versus depth diagrams of oxygen saturation ratio (panels on top) and Chl-a (panels on bottom) as modeled by CMEMS-BGC (panels on the left-hand side) and as emulated by the neural network (NN) when applied to CMEMS physics (panels on the right-hand side) during the period 2015–2019 along the path of the bas1 float.
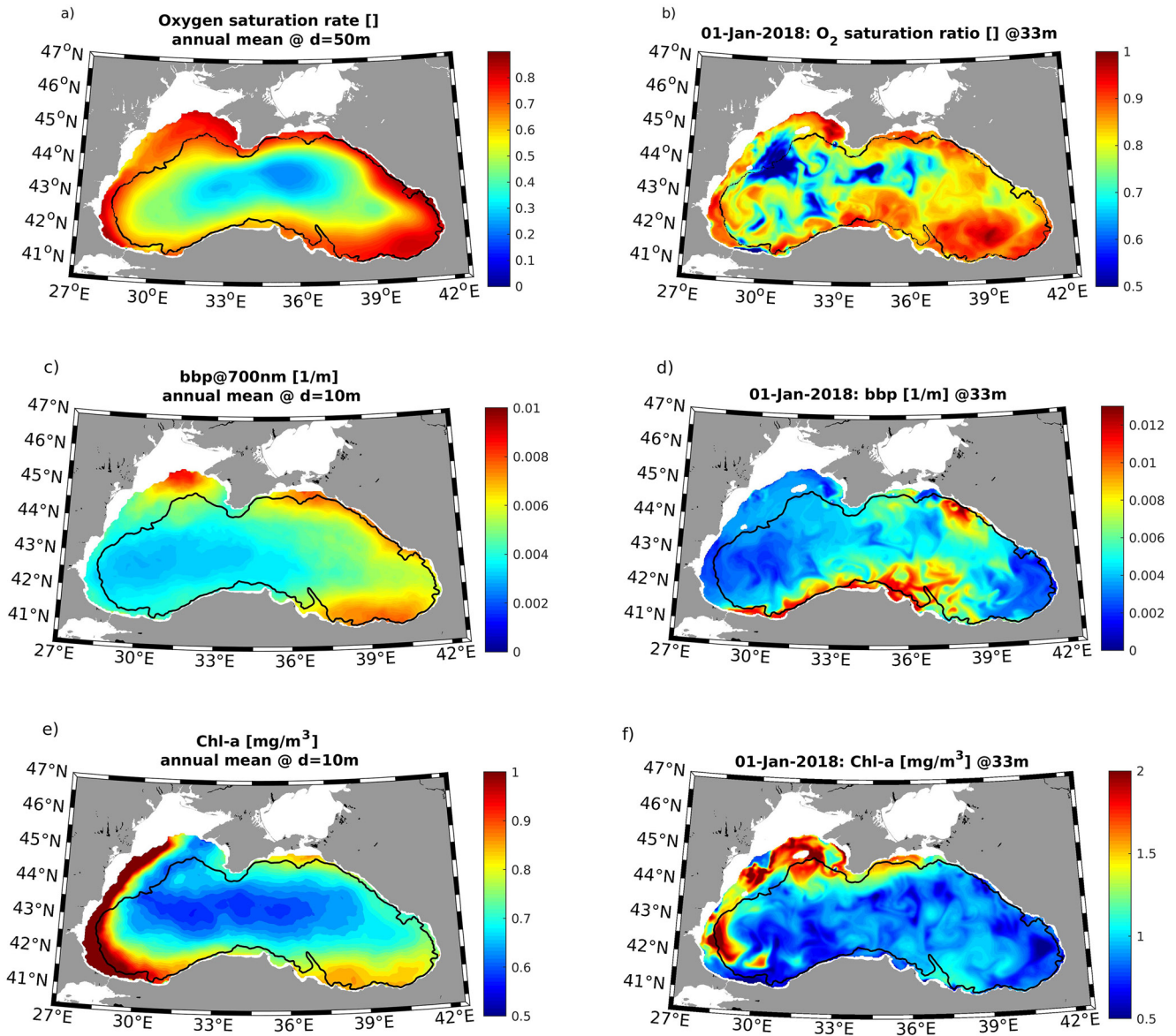
Figure 11. Mean for the 2015–2020 oxygen saturation ratio at 50 m (a), bbp at 10 m (c) and Chl-a at 10 m (e). Panels on the right-hand side show snapshots at 1 January 2018 of oxygen saturation ratio (b), bbp (d) and Chl-a (f) at 33 m. The bathymetric line at 500 m is also plotted to illustrate the deep area, which is usually visited by the profiling floats. The Neural network (NN) is set up to work in areas deeper than 50 m; therefore, the data coverage for all panels is the same.

shows maxima at the end of the winter season, the mean oxygen concentration at 50 m shows a clear decreasing trend (an order of magnitude larger than the accuracy of the optode) overlain by small minima in the early winter. In agreement with Figure 6e, the weaker surface bbp maximum in winter penetrates deeper than the summer maximum over the entire basin. The surface maximum of Chl-a occurs at the same time as when the minimum at 50 m occurs. Furthermore, the temporal variability of Chl-a shows a clear temporal asymmetry: it takes approximately half the time for the transition from Chl-a minimum to Chl-a maximum than for the inverse transition.

The proposed method can also be used to investigate the temporal changes in the spatial patterns of BGC properties. The example below is given for oxygen. The changes in physical state derived from the CMEMS data are shown in Figure 13 for the mean temperature at the top 70 m, sea surface height (SSH), and salinity at 50 m as differences between the states in the periods 2015–2020 (NN training) and 2010–2011 (awi1 and hzg1 float operation). The pattern of the respective difference in oxygen saturation ratio (Figure 13a) is similar to the change in SSH and salinity at 50 m: an increasing (decreasing) trend in SSH (salinity) in the eastern Black Sea and an
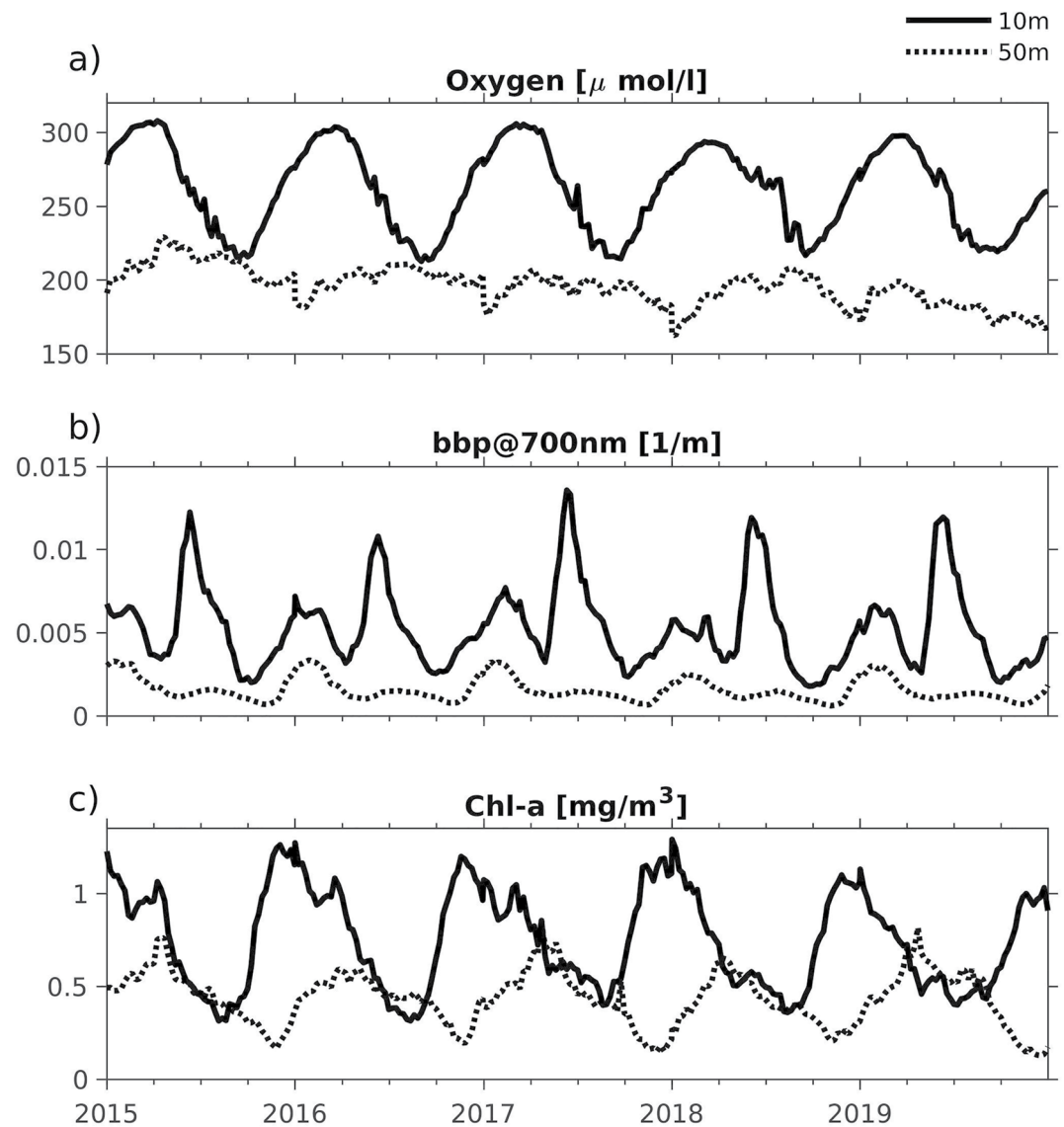
Figure 12. Temporal change in the basin-averaged properties at 10 m (full lines) and 50 m (dotted lines). (a) Oxygen, (b) bbp, (c) Chl-a.

opposite trend in the western Black Sea (Figures 13c and 13d). These trends are large enough to be considered statistically significant (for the accuracy of sensors, see Appendix 1). However, the different and very complex spatial variation in temperature in the upper 70 m (Figure 13b) imposes an additional imprint on oxygen trends. It is notable that the ~8 µmol/l decrease in basin mean oxygen content in the upper 30 m between the two periods is significantly smaller than the ~30 µmol decrease that was estimated from the direct comparison of Argo float readings. We conclude that the existing monitoring platform does not provide a complete description of the state of the entire basin, but with the help of ML and data from physical models, we can overcome the undersampling problem.

4. Discussion

4.1. Mechanisms

Ricour et al. (2021) fitted all the Chl-a profiles from BGC-Argo measurements to five specific mathematical forms with the idea that these profiles should correspond to different states of development of the SCM.
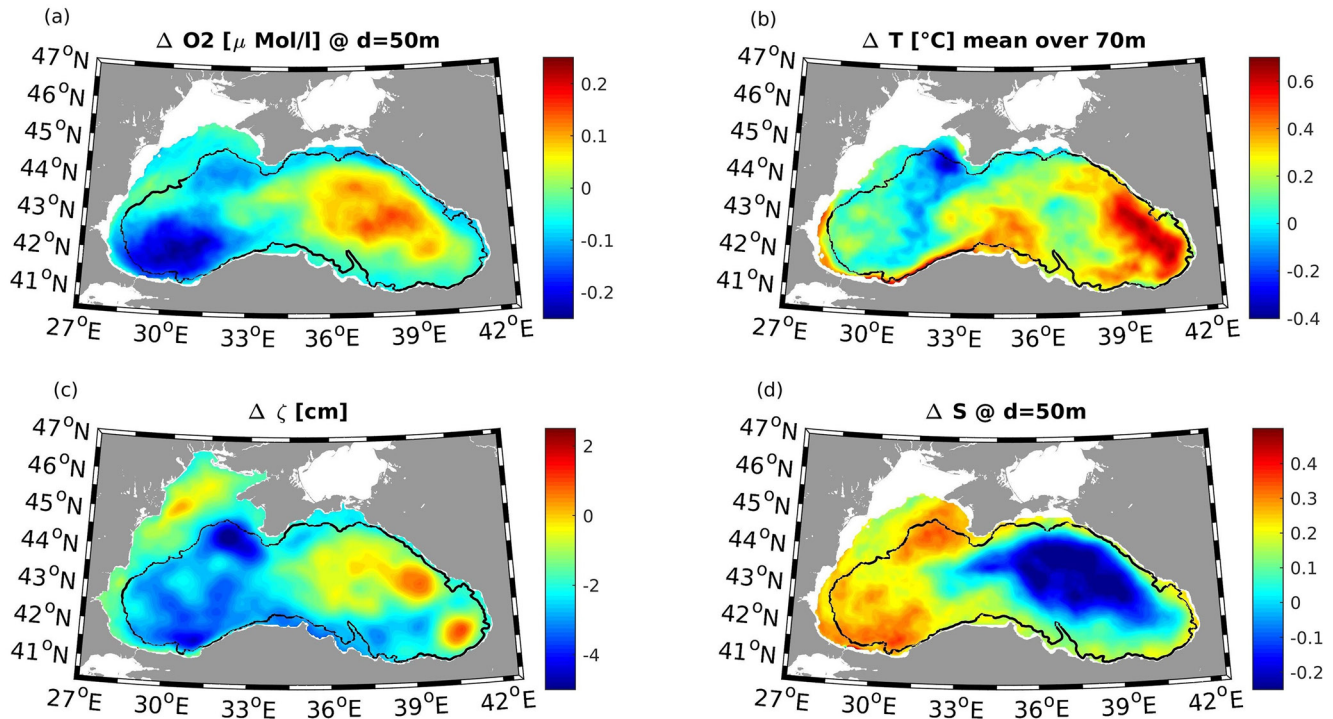
Figure 13. Differences in annual mean values of oxygen saturation ratio (a), mean temperature in the upper 70 m layer (b), sea level height (c), and salinity (d). The average is produced for the periods 2015–2020 (Neural network [NN] training) and 2010–2011 (operation of hzg1 float). Differences are calculated as new minus old period. Oxygen and salinity are shown at a water depth of 50 m.

Kubryakov et al. (2020) defined seven seasonal stages of Chl-a vertical distribution. This "reduction" of the BGC system to a relatively large number of mathematical forms and stages demonstrates that there are no easy explanations for the appearance and causality of the BGC properties in the Black Sea. Our analysis of the NN performance showed that the situation seems highly complex and thus does not allow broad simplifications, which are sometimes used to explain the functioning of BGC fields in the euphotic zone. Furthermore, our analyses showed that the magnitudes of responses of individual BGC parameters to physics appeared not to be proportional to the magnitude of physical variability. This rather complex situation justified using ML as nonlinear data processing.

The analysis of the BGC-Argo observations did not reveal an overshooting of either Chl-a or bbp penetration depth in winter. Such "overshooting patterns" are known from 1-D simulations (Kubryakova et al., 2018; Oguz et al., 1996; Staneva et al., 1998), but did not appear in three-dimensional simulations of Stanev et al. (2014). Moreover, the absence of the "boot heel" pattern in the time versus depth evolution of BGC properties suggests that the mechanism of enrichment of the SCM layer with nitrate-rich waters in realistic BGC systems could be different from what is known from earlier numerical simulations. This finding suggests that this mechanism is essentially not 1-D. More plausibly, nitrate fluxes associated with winter cooling propagate approximately along sloped density surfaces (Stanev et al., 2004). As explained by Cullen (2015), the nutrient flux from below does not need to exclusively originate from vertical mixing. Baroclinicity, which is associated with the slope of isopycnals, can propagate nutrient-rich water (Yentsch, 1980). Therefore, the winter intensification of circulation due to enhanced Sverdrup transport (Stanev et al., 2000), which causes isopycnal surfaces to tilt, could become as important as vertical mixing in modulating the vertical structure of Chl-a. Because of the above specificities, we conclude that "boot-heel-like" patterns are not plausible (at least for the present-day physical state of the Black Sea) and that isopycnal mixing could be as important as vertical convective mixing in winter. The baroclinic mechanism (called a gyre-dominated regime by Stanev et al., 2014) can at least partially explain the "vertical entrainment of nutrients from the nitrocline" proposed by Kubryakova et al. (2021). Unfortunately, the following question remains: did overestimation associated with vertical mixing (the so-called "boot-heel-like" pattern) disappear recently, or was it simply a modeling artifact? Answering this question would necessitate a profound analysis of historical observations, which, unfortunately, are not abundant for the winter season.

Another part of the entrainment of water from the nitrocline is due to the second mixing regime that Stanev et al. (2014) called the eddy-dominated regime in summer. The large differences in the concentrations of Chl-a at stations close to each other indicate small-scale variability. These fluctuations follow the variability of density, which is in line with the hypotheses concerning the important role that eddies play in enhancing the fluxes of nitrate into the euphotic zone (McGillicuddy & Robinson, 1997). Recently, Kubryakova et al. (2021) provided evidence supporting the eddy-dominated exchange regime of coccolithophores in the areas around the continental slope in summer. As demonstrated by Stanev and Staneva (2001), in the Black Sea, mixing associated with mesoscale eddies enhances the penetration of cold surface water into the pycnocline. This process occurs at the periphery of coastal anticyclones. Lateral intrusions of coastal waters further propagate BGC properties into the pycnocline. It seems obvious that the baroclinic mechanism proposed by Yentsch (1980) applies to both gyre-dominated and eddy-dominated regimes of matter exchange. The eddy-like structure of the reconstructed BGC states (panels on the right-hand side of Figure 11) is in line with the above considerations.

4.2. Bbp and Chl-a

Chl-a fluorescence data serve as a proxy for phytoplankton biomass; bbp can be used as a proxy to detect coccolithophore blooms because coccoliths in the Black Sea induce strong backscattering (Kubryakov et al., 2021). Ricour et al. (2021) claimed that bbp is the best proxy that can be obtained from the current BGC-Argo data set in the Black Sea and used this parameter to evaluate the correspondence between chlorophyll and phytoplankton cells. However, the most peculiar result from our analysis is that the seasonal variability of the mean for the euphotic layer Chl-a shows different seasonal variability patterns at different depths, unlike the temperature. Furthermore, the temporal variability of bbp differs largely from that of Chl-a (Figure 12, compare also Figures 1f and 1h)). The spatial patterns of bbp and Chl-a are also very different (Figure 11).

In the Black Sea pycnocline, the bbp at 700 nm is a good proxy for particle concentrations in the size range between the submicroscopic scale and ~10 μm, which includes inorganic particles, heterotrophic bacteria, and coccoliths. Notably, large values of bbp have been detected between the onset of the suboxic zone and the upper part of the anoxic layer (Stanev et al., 2017, 2018), which supports the conclusion of Ediger et al. (2019) that high values of chemoautotrophic biological production appear to be driven by oxygen ventilation and redox cycling. Based on the optical observations presented here, which cannot distinguish between different particles, we cannot speculate on the role of bacteriochlorophyll (Manske et al., 2005; Repeta et al., 1989) in processes in the euphotic zone (remember that the suboxic zone is much deeper than the euphotic layer).

Analyses of satellite and Bio-Argo data by Kubryakova et al. (2021) allowed these authors to suggest that the dynamics of *Emiliania huxleyi* in the Black Sea are strongly affected by intense storms, while their effects on diatoms may not be as pronounced. They also suggested that the response of coccolithophores to extremely high and extremely low irradiances is very different from the response of diatoms. From our analysis, the bbp maxima occur at the sea surface, whereas the Chl-a maximum is deeper (compare Figures 1f and 1h). This observation could indicate a different sensitivity of diatoms and coccolithophores to light conditions.

The comparison between Figures 11c and 11e suggests that one candidate to explain the differences between the horizontal patterns of bbp and Chl-a is river runoff, which is mainly located in the western Black Sea. Thus, one could hypothesize that the influence of freshwater and nutrients from rivers is greater on phytoplankton than on coccolithophores. We also recall that Churilova et al. (2017) explained the difference between bbp and Chl-a, as well as the low correlation between them (see Figure 3b), by the specific eutrophic characteristics of the Black Sea, where a strict relationship between bbp and phytoplankton cell abundance cannot be established. What bbp and Chl-a have in common is that their maximum concentrations are in coastal areas. However, that of Chl-a is in the western basin, while that of bbp is in the eastern basin. Further dedicated studies are needed to better understand the individual and collective dynamics and interdependencies of these two BGC properties. Profiling floats cannot do much more than identify this problem.

4.3. The Undersampling and Trends

Argo BGC data cannot provide a clear view of the basin-wide patterns and temporal evolution of BGC states of the euphotic zone; therefore, extending the interpretation of data from only several floats to the entire basin must be done with care. The basic question in this regard concerns the usefulness of basin-averaged fields based on

limited information. Kubryakov et al. (2020) studied the averaged measurements of four BGC floats in the Black Sea. However, the large spread of data makes simple "basin mean" estimations questionable. Notably, profiles such as those in Figure 6 are mean profiles only for the specific observational platform and must be interpreted with caution. The combined use of an NN with reliable physical information (in our case, from the operational circulation model) could enable reconstructions over the entire basin, even with error estimates. In conclusion, ML helps to address the problem of undersampling by taking advantage of the synergy of different data sources and providing a tool to analyze the 4D structure of BGC fields.

Comparison between the observations used to train the NN and the observations from floats awi1 and hzg1 confirm that the earlier oxygen content in the upper 30 m was higher, which is in line with the results of Capet et al. (2016) regarding the decline of the Black Sea oxygen inventory. A closer examination of the physical data reveals that in 2009–2012, the salinity in the photic layer was lower than that at present (Stanev et al., 2019). This observation led us to explain why we observed a reduction in the reconstruction skill when the NN, trained with data from the last decade, was applied to observations from 2009 to 2012 (Figure 8). This reduction can be considered indirect proof that recent changes in the upper layers of the Black Sea are not just simple trends but are associated with changes in the relationships between the physical and BGC fields. In summary, (a) the decrease in the quality of NN reconstructions when applied to older periods can be used as an instrument to detect changes (trends) in oceanic BGC states, and (b) unfortunately, NNs, which are trained for a relatively short time, are less successful when applied to historical data involving substantial changes in the ocean system.

4.4. Can an NN Be Considered a Substitute for or Complement a Coupled Physical-BGC Model?

There is a common view that NNs can reproduce what has already been observed; therefore, geophysical novelties (new findings) are sometimes not readily clear. We assume that with the presented results, we demonstrate that NNs can do more than that. Analyzing the results from NN experiments can be considered an extension of (linear) correlation analysis, which is usually used to explain the interdependencies between physical and BGC fields and to understand causalities. The better correlations of oxygen with the physical fields, compared with the correlations of bbp and Chl-a, explain the better NN performance for oxygen. The lower skill in the reconstructions of Chl-a could (a) identify excessively complex physical dependencies and (b) indicate that the available data are too sparse to ensure good reconstructions. More plausibly, the dynamics of Chl-a in the Black Sea are less dependent than oxygen dynamics on physics, or the dependencies are more complex.

Notably, the developed NN was not subjected to further improvements to ensure better performance. We admit that with further effort, the quality of the reconstructions could be further improved. However, this was not the aim of the present work, which can be understood as a feasibility study rather than as a technical implementation. The applicability of the developed NN to other ocean areas could also be interesting.

One optimistic result from this study is that the proposed method outperforms the coupled physical-BGC model used in operational services, which could motivate a comparison of the consistency of the NN reconstructions with other coupled models. Provided that the performance of an NN remains superior, NN-produced data could be assimilated into coupled models to improve their skill, which could result in improving the model performance for other BGC state variables.

5. Conclusions

The dynamics of oxygen, bbp and Chl-a obey very different variabilities in space and time. Coccolithophore blooms appear almost synchronously over the entire basin. The SCM layer is characterized by small-scale variability, which, in most cases, follows the variability of density. However, a large portion of the variability in Chl-a concentration is due to BGC processes (not only driven by physics).

To fully account for the complexity in the interrelationships between physical and BGC properties, we developed an NN model using physical data from profiling floats as inputs to predict BGC states. The performance was very high, particularly for oxygen. However, when applying the NN to older data, the performance was reduced, illustrating that the interrelationships between physical and BGC properties in the Black Sea have changed recently. This finding supports the idea that recent climatic changes have affected fundamental BGC mechanisms. Finally, an experiment aiming to reconstruct the BGC states from physical data based on numerical simulations showed,

for the analyzed variables, better performance than the coupled physical-BGC operational model of CMEMS. This finding reveals the potential to enhance the quality of BGC numerical simulations by using ML. One illustration addressed in the paper is the reconstruction of 4D BGC dynamics. The basin-wide temporal variability in Chl-a is strongly asymmetric, and the transition time from the Chl-a minimum to the Chl-a maximum is approximately half that for the inverse transition. Furthermore, we demonstrated that ML fed by physical data from numerical models appears to be a useful tool to replicate both the mean and eddy states of the Black Sea BGC as well as to identify the spatial patterns of recent BGC changes.

Appendix 1: Details About the Observations

Physical variables

- Pressure, temperature and salinity were acquired by an SBE 41 CTD (Sea-Bird Scientific Inc.). The accuracies of temperature, salinity, and pressure sensors are better than 0.002°C, 0.010, and 5 m, respectively (Argo, 2021).

Optical measurements

- Photosynthetically available radiation (PAR) and the downwelling irradiance at 380, 410, and 490 nm were measured using a multispectral radiometer OCR-504 (SAtlantic, Inc.)
- The fluorescence of chlorophyll was measured using a Wetlabs ECO Triplet Puck (WET Labs, Inc.). "CHLA_ADJUSTED" was downloaded from ftp://ftp.ifremer.fr/ifremer/argo. It includes the correction of nonphotochemical quenching (Roesler et al., 2017) as well as the correction of the contribution of fluorescence by nonalgal organic matter (Xing et al., 2017); see also Organelli et al. (2017) and the description of the protocol used at http://www.argodatamgt.org/Documentation.
- bbp(700), particulate backscattering at 700 nm, was measured using a WETLabs ECO Triplet Puck. The measured volume scattering coefficient (m^{-1} sr^{-1}) at 700 nm was converted into bbp(700), as proposed by Schmechtig et al. (2015).
- Oxygen (O_2) was measured using an Aanderra 4330 optode.
- Nitrate (NO_3) was measured using a Satlantic Submersible Ultraviolet Nitrate Analyzer (SUNA) sensor (Johnson & Coletti, 2002; Johnson et al., 2013). According to Rasse et al. (2020), the respective detection limit is ~0.5 μM.

Data with a quality control (QC) flag of 4 ("bad data") were removed (Argo Data Management Team, 2021; Schmechtig et al., 2018), while data with a QCD3 ("probably bad data") were retained. More details on the row data processing can be found in Rasse et al. (2020). The vertical resolution of data acquisition was 10 m between 1,000 and 250 m, 1 m between 250 and 10 m, and 0.2 m between 10 m and the surface.

Profiling floats used iridium transmissions to send the raw data to land.

Some specificities of the Black Sea environment need to be mentioned here. Biofouling is not an issue because floats spend most of their time in dark environments and periodically undergo drastic changes in pressure. In addition, the sulfidic layers of the Black Sea provide another measure of protection against biofouling. Every profile is naturally calibrated with respect to oxygen because a "solid zero" exists at depth; data from the anoxic layers never show oxygen values exceeding 1 μM (Stanev et al., 2017), which can be considered an error estimate for oxygen.

Data Availability Statement

Model data and data from Argo floats used in this study are available at

- https://resources.marine.copernicus.eu/product-detail/BLKSEA_MULTIYEAR_PHY_007_004/INFORMATION
- https://resources.marine.copernicus.eu/product-detail/BLKSEA_REANALYSIS_BIO_007_005/INFORMATION
- http://www.coriolis.eu.org/Data-Products/Data-selection.

Input data as used in this study and BGC states produced by the Neural Network are available at https://zenodo.org/record/6860705#.Yte-CvexXWI

References

Argo (2021). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). *SEANOE*. https://doi.org/10.17882/42182

Barale, V., & Murray, C. N. (1995). The surface colour field of enclosed marine basins: Pigment patterns of the Black Sea. *Remote Sensing Reviews*, *12*(1–2), 61–82. https://doi.org/10.1080/02757259509532276

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Clarendon Press.482.

Capet, A., Stanev, E., Beckers, J. M., Murray, J., & Gregoire, M. (2016). Decline of the Black Sea oxygen inventory. *Biogeosciences*, *13*(4), 1287–1297. https://doi.org/10.5194/bg-13-1287-2016

Churilova, T., Suslin, V., Krivenko, O., Efimova, T., Moiseeva, N., Mukhanov, V., & Smirnova, L. (2017). Light absorption by phytoplankton in the upper mixed layerof the 918 Black Sea: Seasonality and Parametrization. *Frontiers in Marine Science*, *4*, 90.

Ciliberti, S. A., Grégoire, M., Staneva, J., Palazov, A., Coppini, G., Lecci, R., et al. (2021). Monitoring and forecasting the ocean state and biogeochemical processes in the Black Sea: Recent developments in the Copernicus Marine service. *Journal of Marine Science and Engineering*, *9*(10), 1146. https://doi.org/10.3390/jmse9101146

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the global ocean with biogeochemical-Argo. *Annual Review of Marine Science*, *12*(1), 23–48. https://doi.org/10.1146/annurev-marine-010419-010956

Cullen, J. J. (2015). Subsurface chlorophyll maximum layers: Enduring enigma or mystery solved? *Annual Review of Marine Science*, *7*(1), 207–239. https://doi.org/10.1146/annurev-marine-010213-135111

Ediger, D., Murray, J. W., & Yılmaz, A. (2019). Phytoplankton biomass, primary production and chemoautotrophic production of the Western Black Sea in April 2003. *Journal of Marine Systems*, *198*, 103183. https://doi.org/10.1016/j.jmarsys.2019.103183

Ginsburg, A. I., Kostianoy, A. G., Soloviev, D. M., & Stanichny, S. V. (1997). Coastal upwelling in the north-west Black Sea. *Issled Zemli Kosm*, *6*, 61–72. (in Russian).

Grégoire, M., Raick, C., & Soetaert, K. (2008). Numerical modeling of the deep Black Sea ecosystem functioning during the late 80's (eutrophication phase). *Progress in Oceanography*, *76*(3), 286–333. https://doi.org/10.1016/j.pocean.2008.01.002

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, *4*(2), 251–257. https://doi.org/10.1016/0893-6080(91)90009-t

Johnson, K., Coletti, L., Jannasch, H., Sakamoto, C., Swift, D., & Riser, S. (2013). Long-term nitrate measurements in the ocean using the in situ ultraviolet spectrophotometer: Sensor integration into the Apex profiling float. *Journal of Atmospheric and Oceanic Technology*, *30*(8), 1854–1866. https://doi.org/10.1175/JTECH-D-12-00221.1

Johnson, K. S., & Coletti, L. J. (2002). In situ ultraviolet spectrophotometry for high resolution and long-term monitoring of nitrate, bromide and bisulfide in the ocean. *Deep-Sea Research Part I*, *49*(7), 1291–1305. https://doi.org/10.1016/S0967-0637(02)00020-1

Kara, B. A., Helber, R. W., Boyer, T. P., & Elsner, J. B. (2009). Mixed layer depth in the Aegean, Marmara, Black and Azov Seas: Part I: General features. *Journal of Marine Systems*, *78*(suppl), S169–S180. https://doi.org/10.1016/j.jmarsys.2009.01.022

Karabashev, G. S., Evdoshenko, M. A., & Sheberstov, S. V. (2006). Normalized radiance spectrum as a water exchange event diagnostic. *International Journal of Remote Sensing*, *27*(9), 1775–1792. https://doi.org/10.1080/01431160500380505

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, 107398. https://doi.org/10.1016/j.ymssp.2020.107398

Konovalov, S., & Murray, J. (2001). Variations in the chemistry of the Black Sea on a time scale of decades (1960–1995). *Journal of Marine System*, *31*(1–3), 217–243. https://doi.org/10.1016/s0924-7963(01)00054-9

Konovalov, S. K., Murray, J. W., & Luther, G. W., III. (2005). Basic processes of Black Sea biogeochemistry. *Oceanography*, *18*(2), 24–35. https://doi.org/10.5670/oceanog.2005.39

Kopelevich, O. V., Burenkov, V. I., Ershova, S. V., Sheberstov, S., & Evdoshenko, M. (2004). Application of SeaWiFS data for studying of bio-optical characteristics in the Barents, Black and Caspian Seas. *Deep-Sea Research II*, *51*(10–11), 1063–1091. https://doi.org/10.1016/s0967-0645(04)00101-8

Kubryakov, A. A., Belokopytov, V. N., Zatsepin, A. G., Stanichny, S. V., & Piotukh, V. B. (2019a). The Black Sea mixed layer depth variability and its relation to the basin dynamics and atmospheric forcing. *Physical Oceanography*, *26*(5), 397. https://doi.org/10.22449/1573-160X-2019-5-397-413

Kubryakov, A. A., Mikaelyan, A. S., & Stanichny, S. V. (2019). Summer and winter coccolithophore blooms in the Black Sea and their impact on production of dissolved organic matter from Bio-Argo data. *Journal of Marine Systems*, *199*, 103220. https://doi.org/10.1016/j.jmarsys.2019.103220

Kubryakov, A. A., Mikaelyan, A. S., & Stanichny, S. V. (2021). Extremely strong coccolithophore blooms in the Black Sea: The decisive role of winter vertical entrainment of deep water. *Deep Sea Research Part I: Oceanographic Research Papers*, *173*, 103554. https://doi.org/10.1016/j.dsr.2021.103554

Kubryakov, A. A., Mikaelyan, A. S., Stanichny, S. V., & Kubryakova, E. A. (2020). Seasonal stages of chlorophyll-a vertical distribution and its relation to the light conditions in the Black Sea from Bio-Argo measurements. *Journal of Geophysical Research: Oceans*, *125*(12), e2020JC016790. https://doi.org/10.1029/2020JC016790

Kubryakova, E. A., Kubryakov, A. A., & Mikaelyan, A. S. (2021). Winter coccolithophore blooms in the Black Sea: Interannual variability and driving factors. *Journal of Marine Systems*, *213*, 103461. https://doi.org/10.1016/j.jmarsys.2020.103461

Kubryakova, E. A., & Kubryakov, A. A., & Stanichny, S. V. (2018). Impact of winter cooling on water vertical entrainment and intensity of phytoplankton bloom in the Black Sea. *Physical Oceanography*, *25*(3), 191–206. https://doi.org/10.22449/1573-160X-2018-3-191-206

Madec, G., & the NEMO team. (2016). *NEMO ocean engine, Note du Pôle de modélisation* (pp. 1288–1619). Institut Pierre-Simon Laplace No 27.

Manske, A. K., Glaeser, J., Kuypers, M. M. M., & Overmann, J. (2005). Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. *Applied and Environmental Microbiology*, *71*(12), 8049–8060. https://doi.org/10.1128/aem.71.12.8049-8060.2005

McGillicuddy, D. J., Jr., & Robinson, A. R. (1997). Eddy-induced nutrient supply and new production in the Sargasso Sea. *Deep-Sea Research I*, *44*(8), 1427–1450. https://doi.org/10.1016/s0967-0637(97)00024-1

Mikaelyan, A. S., Shapiro, G. I., Chasovnikov, V. K., Wobus, F., & Zanacchi, M. (2017). Drivers of the autumn phytoplankton development in the open Black Sea. *Journal of Marine Systems*, *174*, 1–11. https://doi.org/10.1016/j.jmarsys.2017.05.006

Murray, J. W., Jannasch, H. W., Honjo, S., Anderson, R. F., Reeburgh, W. S., Top, Z., et al. (1989). Unexpected changes in the oxic/anoxic interface in the Black Sea. *Nature*, *338*(6214), 411–413. https://doi.org/10.1038/338411a0

Murray, J. W., & Yakushev, E. V. (2006). The suboxic transition zone in the Black Sea. In L. N. Neretin (Ed.), *Past and present water column anoxia. NATO Sciences Series* (pp. 105–138). Springer. https://doi.org/10.1007/1-4020-4297-3_05

Oguz, T., Ducklow, H. W., Malanotte-Rizzoli, P., Tugrul, S., Nezlin, N. P., & Unluata, U. (1996). Simulation of annual plankton productivity cycle in the Black Sea by a one-dimensionalphysical-biological model. *Journal of Geophysical Research*, *101*(C7), 16585–16599. https://doi.org/10.1029/96jc00831

Organelli, E., Barbieux, M., Claustre, H., Schmechtig, C., Poteau, A., Bricaud, A., et al. (2017). Two databases derived from BGC-Argo float measurements for marine biogeochemical and bio-optical applications. *Earth System Science Data*, *9*, 861–880. https://doi.org/10.5194/essd-9-861-2017

Rasse, R., Claustre, H., & Poteau, A. (2020). The suspended small-particle layer in the oxygen-poor Black Sea: A proxy for delineating the effective N₂-yielding section. *Biogeosciences*, *17*(24), 6491–6505. https://doi.org/10.5194/bg-17-6491-2020

Repeta, D. J., Simpson, D. J., Jorgensen, B. B., & Jannasch, H. W. (1989). Evidence for anoxygenic photosynthesis from the distribution of bacteriochlorophylls in the Black Sea. *Nature*, *342*(6245), 69–72. https://doi.org/10.1038/342069a0

Richardson, A. J., Pfaff, M. C., Field, J. G., Silulwane, N. F., & Shillington, F. A. (2002). Identifying characteristic chlorophyll aprofiles in the coastal domain using an artificial neural network. *Journal of Plankton Research*, *24*(12), 1289–1303. https://doi.org/10.1093/plankt/24.12.1289

Ricour, F., Capet, A., D'Ortenzio, F., Delille, B., & Grégoire, M. (2021). Dynamics of the deep chlorophyll maximum in the Black Sea as depicted by BGC-Argo floats. *Biogeosciences*, *18*(2), 755–774. https://doi.org/10.5194/bg-18-755-2021

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., et al. (2017). Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. *Limnology and Oceanography: Methods*, *15*(6), 572–585. https://doi.org/10.1002/lom3.10185

Sammartino, M., Buongiorno Nardelli, B., Marullo, S., & Santoleri, R. (2020). An artificial neural network to infer the Mediterranean 3D chlorophyll-a and temperature fields from remote sensing observations. *Remote Sensing*, *12*(24), 4123. https://doi.org/10.3390/rs12244123

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., & Johnson, K. S. (2017). Estimates of Water-column nutrient concentrations and carbonate system parameters in the Global Ocean: A novel approach based on neural networks. *Frontiers in Marine Science*, *4*, 128. https://doi.org/10.3389/fmars.2017.00128

Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., & D'Ortenzio, F. (2015). Retrieving the vertical distribution of chlorophyll a concentration andphytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications. *Journal Geophysical Research: Ocean*, *120*(1), 451–470. https://doi.org/10.1002/2014JC010355

Schiller, H. (2000). Feedforward-backpropagation neural net program ffbp1.0. GKSSreport 2000/37, ISSN 0344-9629.

Schiller, H., & Doerffer, R. (1999). Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing*, *20*(9), 1735–1746. https://doi.org/10.1080/014311699212443

Schmechtig, C., Claustre, H., Poteau, A., & D'Ortenzio, F. (2018). Bio-Argo quality control manual for the Chlorophyll-A concentration. *Ifremer*. https://doi.org/10.13155/35385

Schmechtig, C., Poteau, A., Claustre, H., D'Ortenzio, F., & Boss, E. (2015). Processing bio-Argo chlorophyll a concentration at the DAC level. *Argo Data Management*, 1–22. https://doi.org/10.13155/39468

Stanev, E., Le Traon, P., & Peneva, E. (2000). Sea level variations and their dependency on meteorological and hydrological forcing: Analysis of altimeter and surface data for the Black Sea. *Journal of Geophysical Research*, *105*(170), 17203–17216. https://doi.org/10.1029/1999jc900318

Stanev, E. V., & Chtirkova, B. (2021). Interannual change in mode waters: Case of the black sea. *Journal of Geophysical Research: Oceans*, *126*(2), e2020JC016429. https://doi.org/10.1029/2020JC016429

Stanev, E. V., Grayek, S., SchmechtigPoteau, H. C. A., Schmechtig, C., & Poteau, A. (2017). Water intrusions and particle signatures in the Black Sea: A biogeochemical-Argo float investigation. *Ocean Dynamics*, *67*(9), 1119–1136. https://doi.org/10.1007/s10236-017-1077-9

Stanev, E. V., He, Y., Grayek, S., & Boetius, A. (2013). Oxygen dynamics in the Black Sea as seen by Argo profiling floats. *Geophysical Research Letters*, *40*(12), 3085–3090. https://doi.org/10.1002/grl.50606

Stanev, E. V., He, Y., Staneva, J., & Yakushev, E. (2014). Mixing in the Black Sea detected from the temporal and spatial variability of oxygen and sulfide—Argo float observations and numerical modelling. *Biogeosciences*, *11*(20), 5707–5732. https://doi.org/10.5194/bg-11-5707-2014

Stanev, E. V., Peneva, E., & Chtirkova, B. (2019). Climate change and regional ocean water mass disappearance: Case of the Black Sea. *Journal of Geophysical Research: Oceans*, *124*(7), 4803–4819. https://doi.org/10.1029/2019JC015076

Stanev, E. V., Poulain, P.-M., Grayek, S., Johnson, K. S., Claustre, H., & Murray, J. W. (2018). Understanding the dynamics of the oxic-anoxic interface in the Black Sea. *Geophysical Research Letters*, *45*(2), 864–871. https://doi.org/10.1002/2017GL076206

Stanev, E. V., Staneva, J., Bullister, J. L., & Murray, J. W. (2004). Ventilation of the Black Sea pycnocline. Parameterization of convection, numerical simulations and validations against observed chlorofluorocarbon data. *Deep-Sea Research*, *51/12*(12), 2137–2169. https://doi.org/10.1016/j.dsr.2004.07.018

Stanev, E. V., & Staneva, J. V. (2001). The sensitivity of the heat exchange at sea surface to meso and sub-basin scale eddies. Model study for the Black Sea. *Dynamics of Atmospheres and Oceans*, *33*, 163–189.

Staneva, J. V., Stanev, E. V., & Oguz, T. (1998). The impact of atmospheric forcing and water column stratification on the yearly plankton cycle. In L. Ivanov & T. Oguz (Eds.), *Ecosystem modelling as a management tool for the Black Sea* (Vol. 2, pp. 301–322). Kluwer Academic Publishers.

Sur, H. I., Ozsoy, E., Ilyin, Y. P., & Unluata, U. (1996). Coastal-deep ocean interactions in the Black Sea and their ecological/environmental impacts. *Journal of Marine Systems*, *7*(2–4), 293–320. https://doi.org/10.1016/0924-7963(95)00030-5

Tugrul, S., Basturk, O., Saydam, C., & Yilmaz, A. (1992). Changes in the hydrochemistry of the Black Sea inferred from water density profiles. *Nature*, *359*(6391), 137–139. https://doi.org/10.1038/359137a0

Willmott, C. J. (1981). On the validation of models. *Physical Geography*, *2*(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213

Xing, X., Claustre, H., Boss, E., Roesler, C., Organelli, E., Poteau, A., et al. (2017). Correction of profiles of in-situ chlorophyll fluorometry for the contribution of fluorescence originating from non-algal matter. *Limnology and Oceanography: Methods*, *15*(1), 80–93. https://doi.org/10.1002/lom3.10144

Yakushev, E. V., Chasovnikov, V. K., Murray, J. W., Pakhomova, S. V., Podymov, O. I., & Stunzhas, P. A. (2007). Vertical hydrochemical structure of the Black Sea. In: *The Black Sea environment* (pp. 277–307). Springer.

Yentsch, C. S. (1980). Phytoplankton growth in the sea: A coalescence of disciplines. In P. G. Falkowski (Ed.), *Primary productivity in the sea* (pp. 17–31). Plenum.