

Estimating Seasonal Global Sea Surface Chlorophyll-a with ~~Resource-Efficient Neural Networks~~ a Lightweight Data-Driven Approach

Gabriela Martinez Balbontin¹, Julien Jouanno^{2,3}, Rachid Benshila², Julien Lamouroux¹, Coralie Perruche¹, and Stefano Ciavatta¹

¹Mercator Océan International, Toulouse, France

²Laboratoire d'Etudes en Géophysique et Océanographie Spatiales, Toulouse, France

³Institut de Recherche pour le Développement, France

Correspondence: Gabriela Martinez Balbontin (gmartinezbalbontin@mercator-ocean.fr)

Abstract. Marine chlorophyll-a ~~is an important concentration is a key~~ indicator of ecosystem health, and accurate ~~forecasting, even at the surface level, can have significant implications for climate studies and resource management. Traditionally, these predictions have relied~~ seasonal forecasting has important applications for fisheries management, harmful algal bloom detection, and climate studies. ~~Traditional approaches rely~~ on computationally intensive numerical ~~models, which biogeochemical models~~ ~~that~~ require extensive domain expertise and ~~careful~~ parameterization.

We propose a data-driven, ~~resource-efficient~~ alternative: a ~~lightweight, resource-efficient~~ neural architecture based on the U-Net that reconstructs surface, near-global chlorophyll-a ~~from~~ based on observations and four physical predictors. The model ~~uses~~ learns to emulate satellite-like chlorophyll-a from mixed layer depth, sea surface ~~temperature, sea surface height~~, salinity, and ~~sea surface height temperature~~ as input, all of which are known to influence phytoplankton distribution and nutrient availability. By leveraging publicly available seasonal forecasts of these variables, we can generate six-month of chlorophyll-a predictions in a matter of minutes with a single GPU.

We ~~first validated the quality of the reconstruction by trained the model~~ using the GLORYS12 reanalysis as input. ~~The reconstructed time series demonstrated strong agreement with the reference GlobColour observations, with an RMSE of 0.01 and a correlation of 0.95. Extending this approach~~ physics and GlobColour merged chlorophyll-a observations as reference. ~~When applied~~ to seasonal forecasting, ~~we used six-month by using SEAS5 ensemble forecasts as input and found that our predictions maintained as opposed to the reanalysis, the model maintains~~ high skill globally, ~~with low error rates and stable correlation coefficients throughout the forecast period.~~

~~Our~~ and remains stable across the six-month forecast horizon. Regional analysis shows that the model accurately captures ~~spatial and temporal~~ seasonal dynamics and bloom timings across diverse regimes, with performance comparable to or ~~exceeding that of a state-of-the-art numerical biogeochemical model, while requiring orders of magnitude less computational resources.~~

This approach demonstrates that high-performing, seasonal chlorophyll-a ~~patterns across a variety of regions, with an accuracy that meets or exceeds that of the numerical model of reference while significantly reducing computational costs. This~~

~~approach offers a scalable, efficient alternative for long-term chlorophyll-a forecasting~~forecasting can be achieved through a resource-efficient, observation-driven framework, offering a practical alternative for operational applications where computational constraints limit the use of full biogeochemical models.

1 Introduction

Like many other natural systems, marine biogeochemical cycles are vulnerable to the effects of climate change (Bopp et al., 2013; Achterberg, 2014). Stressors such as rising temperatures (Polovina et al., 2011), ocean acidification (Orr et al., 2005), and changes in nutrient availability (Achterberg, 2014; Schneider et al., 2008) are already affecting the oceans, and these impacts are expected to intensify in the coming years.

These changes can have cascading effects on marine ecosystems, causing alterations in species distribution and biodiversity, impacting primary production, and posing risks to food security. An important indicator of changes in the health of aquatic ecosystems is chlorophyll-a (chl-a) concentration. This photosynthetic pigment can be used to estimate phytoplankton biomass and productivity, and it has the additional advantage of being observable from space (Groom et al., 2019), which allows for its large-scale monitoring.

Accurate seasonal chl-a forecasting, even if limited to the surface level, can have a variety of important applications. It can help estimate food availability for higher trophic levels, with implications for fisheries management and biodiversity conservation (Park et al., 2019b). Additionally, it enables the early detection of harmful algal blooms, allowing for timely warnings to mitigate their impacts on human health and coastal economies (Lin et al., 2021; Wang et al., 2018). Chl-a forecasting also serves as a valuable indicator of upwelling events, which influence nutrient availability and primary production. These processes, in turn, have effects on the carbon cycle and the broader climate system (Bopp et al., 2013; Achterberg, 2014; Rousseaux et al., 2021; Park et al., 2019b).

Traditionally, biogeochemical forecasting has been done using numerical models, which are based on differential equations that model nutrient and carbon cycles (Aumont et al., 2015; Fennel et al., 2022; Berardi, 2020; Gehlen et al., 2015). While some of these models have successfully been used for seasonal chl-a forecasting (Rousseaux et al., 2021; Park et al., 2019b), they can be very computationally prohibitive, and they require extensive knowledge of the systems they represent. Incorrect parametrization can lead to systematic biases and uncertainties, and these errors can propagate throughout the model, limiting its capabilities (Fennel et al., 2022; Berardi, 2020; Gehlen et al., 2015).

Machine learning offers the possibility of making use of the growing number of geospatial data for this purpose. While its use in marine biogeochemical modeling is not as widespread as in meteorology, its adoption is steadily growing (Sadaiappan et al., 2023). In the case of chl-a, it has been used to enhance remote sensing datasets by filling observational gaps and correcting biases caused by sensor limitations or cloud cover (Keiner, 1999; Hu et al., 2021; Kolluru and Tiwari, 2022; Park et al., 2019a; Cao et al., 2020). Martinez et al. (2020) reconstructed historical chl-a patterns from surface oceanic and atmospheric ~~parameters~~predictors, demonstrating that this approach is able to capture decadal and longer trends of past periods, while Rousillon et al. (2023) further expanded on this work, highlighting the different regional modes of variability. Machine learning

has also been applied to short-term chl-a forecasting, primarily at regional scales. Park et al. (2015) proposed a support vector machine for forecasts within freshwater and estuarine reservoirs, while Wenxiang et al. (2022) used a neural network to generate forecasts in the Xiamen Bay. Additional examples include Chen et al. (2024), who generated forecasts for freshwater lakes, and Ly et al. (2021), who focused on the Han River. Zhu et al. (2023) and Ying et al. (2023) applied data-driven methods for forecasting in coastal waters. Nonetheless, to our knowledge, most existing efforts focus on regional-scale forecasts spanning only a few days.

Many of these studies demonstrate the feasibility of reconstructing chl-a using physical data (Martinez et al., 2020; Roussillon et al., 2022; Martinez et al., 2020; Roussillon et al., 2023; Chen et al., 2024; Sauzède et al., 2017). Based on this, we propose a machine learning-driven methodology ~~that can predict to predict six-months of~~ global, surface-level chl-a ~~given by using forecasts for the same time period of~~ four oceanic variables: mixed layer depth (MLD), sea surface temperature (SST), sea surface salinity (SSS), and sea surface height (SSH) ~~at a spatial resolution of approximately 25 km (1/4 degree) and a temporal resolution of 5-days to a month.~~ These variables are ~~known to influence~~ selected due to their influence on nutrient and oxygen availability, ~~and as well as~~ phytoplankton distribution, making them ~~good-effective~~ predictors of chl-a (Browning and Moore, 2023; Palacios et al., 2013; Fernandez-Gonzalez et al., 2022; Xu et al., 2022; Barone et al., 2019; Uz et al., 2001; Chenillat et al., 2021). They also offer practical advantages: they are readily available from ocean reanalyses, and it is possible to obtain their seasonal forecasts.

The goal of this work is to demonstrate that we can not only estimate chl-a from these four variables, but that by using publicly available forecasts of these as input, we are able to generate an ensemble of ~~skillful-accurate~~ chl-a predictions for six months into the future. Our approach operates at a spatial resolution of approximately 25 km (1/4°) and two different temporal resolutions: 5-days (as a proof-of-concept of performance at higher temporal resolution) and monthly (to match the available physics forecasts used as input).

2 Proposed Methodology

2.1 ~~Model architecture and training data~~

The predictive task consists of reconstructing the chl-a distribution over a six-month horizon using forecasts of four physical predictors. Given the intended practical utility of these predictions in operational settings, our primary objective was to develop a solution that is computationally efficient, operationally robust, and readily deployable under limited computational resources.

Because the target exhibits strong seasonality and our focus was not necessarily maximal architectural expressiveness, we did not adopt a single monolithic model covering the full range of variability. Instead, we implemented a month-dependent modeling strategy composed of twelve lightweight neural networks, each corresponding to the initialization month of the forecast. The training period spans January 1998 to May 2017, and each network was trained on six-month time series starting from its initialization month (m_1 - m_6). For example, the network initialized in January was trained on January-June data from

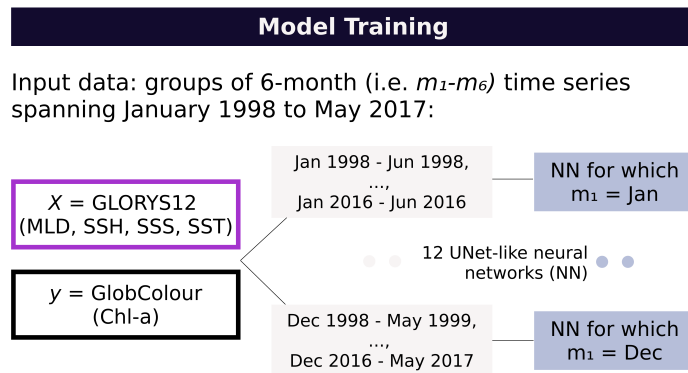


Figure 1. Overview of the methodology: Twelve month-specific neural networks are trained on six-month time series spanning 1998-2017, each initialized on the starting month of the forecast.

1998-2016, while the December network was trained on December 1998-May 2017, and so on. Figure 1 provides an overview of the methodology.

This strategy limits the framework’s ability to generalize across initialization months and may reduce capacity to learn universal temporal relationships. Nonetheless, it offers some practical advantages: each network is substantially smaller than a unified model, conditioning on the initialization month reduces representational complexity, and the design mirrors the operational workflow, in which users generate chl-a forecasts based on physics predictions available from a specific month of initialization. This approach also allows each model to focus on a narrower distribution of patterns, which can be learned effectively with fewer parameters; thus, we considered this an acceptable trade-off.

Two versions of this workflow were created: one using monthly resolution (to match the available physics forecasts that would be used as input operationally), and another using 5-day averages (i.e. 6 data points per month, to test performance at higher temporal resolution). Both architectures use a spatial resolution of $1/4^\circ$, or approximately 25 km.

2.1 Training data

The physics data were provided by the Global Ocean Physics Reanalysis (GLORYS12) (Lellouche et al., 2021; Copernicus Marine Service), and the Global Ocean Color (GlobColour) dataset (Garneison et al., 2019) was used as the output target. GLORYS12 is a high-resolution reanalysis that provides daily and monthly data on the four predictors (MLD, SST, SSS, SSH), among other variables. It is based on the NEMO (Nucleus for European Modelling of the Ocean) platform and driven at the surface by ECMWF’s ERA-Interim and ERA5 meteorological reanalyses. Satellite altimetry, SST, sea ice concentration, and in situ temperature and salinity vertical profiles are assimilated (Lellouche et al., 2021).

The GlobColour dataset provides a combination of gap-filled satellite observations of chl-a from multiple remote sensors. Chl-a concentrations are derived from reflectance using sensor-specific algorithm coefficients and then merged (Garneison et al., 2019; Mar

110 . Because the availability and quality of remotely sensed chl-a degrade at high latitudes due to low solar elevation, persistent cloud cover, and sea-ice, we restrict the domain to a -60° to 80° latitude grid in order to ensure robust training targets.

Within the development dataset (Jan 1998-May 2017), the years 2001, 2005, 2010 and 2015 were used for validation, while the rest was used for training. The training set was used to fit the model, while the validation set served for hyperparameter optimization and selecting the most effective model architecture. After selecting the final parameters using the training and validation sets, the model is evaluated on an independent test set, consisting of the actual forecast fields (2018-2023 for the 115 5-day model and 2020-2023 for the monthly model).

The physical ocean data were normalized using min-max scaling and the chl-a data were log-transformed. Both datasets are available at a daily resolution, and the data were averaged to monthly and 5-day resolutions for training the respective versions of the model. The $1/12^\circ$ GLORYS12 grid was coarsened using bicubic interpolation to match the observational grid resolution, which was obtained at $1/4^\circ$.

120 2.2 Training objective

The proposed methodology reconstructs a six-month chl-a grid (\hat{y}) given ~~an~~ a corresponding six-month input grid of four physical variables (X , i.e. MLD, SST, SSS, SSH). ~~The model is based on the~~ $X = \{MLD, SST, SSS, SSH\}$. As mentioned above, the network operates on log-transformed chl-a values, where $p_i = \log(\hat{y}_i)$ denotes the predicted values for the i -th grid point, and $r_i = \log(y_i)$ denotes the corresponding log-transformed reference point. The training objective is defined as the 125 mean squared error (MSE) between the network output and the reference:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (r_i - p_i)^2 \quad (1)$$

where N is the total number predicted points in the training batch.

In spite of the log-transformation, initial assessments showed that the model had a tendency to underestimate chl-a concentrations, so the standard MSE loss function was modified by adding a small penalty for underestimation. This adjustment was based on 130 empirical testing and was implemented by applying a small weight w whenever the underestimation occurred:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left[(r_i - p_i)^2 + w \cdot [\max(0, r_i - p_i)]^2 \right] \quad (2)$$

where the second term applies an additional penalty $w \cdot (r_i - p_i)^2$ when $p_i < r_i$. While this approach was effective in reducing underestimation, more principled methods could provide a more robust solution.

2.3 Model architecture

135 The neural networks are based on the U-Net architecture ~~, which is~~ (Ronneberger et al., 2015), a convolutional neural network with an encoder-decoder structure that ~~was introduced by Ronneberger et al. (2015). Because of~~ has been widely applied

to Earth science forecasting tasks due to the grid-based, image-like nature of Earth science data, the nature of such data (Agrawal et al., 2019; Sønderby et al., 2020; Weyn et al., 2019, 2021). The U-Net and other variations of convolutional neural networks were among the first deep learning architectures applied to tasks such as weather forecasting (Agrawal et al., 2019; Sønderby et al., 2020). While advances in computational power and data availability have enabled the development of more complex skip connections, enable it to combine low-level spatial details with high-level semantic information, allowing reconstruction of both large-scale patterns and fine-scale features simultaneously. Despite advances and increased complexity in more recent architectures, the U-Net's simplicity and efficiency of the U-Net makes make it an effective and resource-efficient choice for this task.

To process both spatial and temporal dimensions simultaneously, we use 3D convolutions (Tran et al., 2015). To balance the additional computational cost of the 3D convolutions, the proposed architecture uses fewer layers and filters than the original U-Net.

The encoder consists of two blocks of convolutional and max-pooling layers, which downsample the input. Two additional convolutional layers refine feature extraction before transitioning to the decoder, which mirrors the encoder and upsamples the data back to the original resolution. Skip connections link matching layers of corresponding size in the encoder and decoder, facilitating the transfer of information (see Table 1). A rectified linear unit (ReLU) activation function was used throughout the network, except in the output layer, where a Softplus activation function (Dugas et al., 2000) was empirically found to work best. The model architecture is summarized in Table 1.

All twelve models share a common architecture and hyperparameters (i.e. number of layers and filters, and order of layers, number of convolutional filters and kernel size, learning rate, etc.) were optimized, which were optimized empirically and using random search and the model was. The models were trained using the Adam optimizer (Kingma and Ba, 2017). Empirical tests showed that the model had a tendency to underestimate chl-a concentrations, so we modified the standard mean squared error (MSE) loss function by adding a small penalty for underestimation. This adjustment was based on empirical testing and was implemented by applying a weight w , where $0.5 > w > 0.3$, whenever the predicted values p_i were smaller than the reference r_i . While this approach was effective in reducing underestimation, more principled methods could provide a more robust solution. The physical ocean data was normalized using min-max normalization and the chl-a data was log-transformed.

$$\text{Modified MSE loss function} = \frac{1}{N} \sum [(r_i - p_i)^2 + \max(0, r_i - p_i)^2 \cdot w]$$

To simplify the predictive task, which consists of using a 6-month forecast of the physics to predict chl-a for the same time period, we trained twelve dedicated neural networks, each corresponding to the starting month of the forecast. These networks share a common architecture and fixed parameters. Additionally, the trained weights from any one of the twelve networks can be used for and a learning rate of 0.001 for a maximum of 300 epochs. We used early stopping on the validation loss to prevent overfitting, but no explicit regularization methods were used. We also used transfer learning, where the training process of a model can be initialized with the pre-trained weights of another, optimizing to further optimize computational cost.

Table 1. Summary of proposed neural architecture

Layer	Number of trainable parameters	Connected to
Input layer(input_1)	0	-
Conv3D (conv3d_1)	6,976	input_1 Input layer
Conv3D (conv3d_12)	110,656	conv3d_1 Conv3D (1)
MaxPooling3D (max_pooling3d_1)	0	conv3d_1 Conv3D (2)
Conv3D (conv3d_23)	221,312	max_pooling3d_1 MaxPooling3D (1)
Conv3D (conv3d_34)	442,496	conv3d_2 Conv3D (3)
MaxPooling3D (max_pooling3d_12)	0	conv3d_3 Conv3D (4)
Conv3D (conv3d_45)	884,992	max_pooling3d_1 MaxPooling3D (2)
Conv3D (conv3d_56)	1,769,728	conv3d_4 Conv3D (5)
UpSampling3D (up_sampling3d_1)	0	conv3d_5 Conv3D (6)
Concatenation-Skip connection (concatenate_1)	0	up_sampling3d_1 Concat. UpSampling3D (1), conv3d_3 Conv3D (3)
Conv3D (conv3d_67)	1,327,232	concatenate_1 Skip connection (1)
Conv3D (conv3d_78)	442,496	conv3d_6 Conv3D (7)
UpSampling3D (up_sampling3d_12)	0	conv3d_7 Conv3D (8)
Concatenation-Skip connection (concatenate_12)	0	up_sampling3d_1 Concat. UpSampling3D (2), conv3d_1 Conv3D (1)
Conv3D (conv3d_89)	331,840	concatenate_1 Skip connection (2)
Output(conv3d_9)	1,729	conv3d_8 Conv3D (9)

The optimal architecture found for this task has approximately six million trainable parameters. For context, according to some estimates (Shah et al., 2022; Ansari et al., 2022), the original 2D U-Net proposed by Ronneberger et al. has between 20 and 30 million. The training was carried out [The models were trained](#) on a single GPU (NVIDIA A100, 40GB) and took approximately 3-5 hours to complete.

The model was trained using 6-month time series from 1998 to 2016, with physical ocean data from the Global Ocean Physics Reanalysis (GLORYS12) (Lellouche et al., 2021; Copernicus Marine Service, 2024), and chl-a observations from the Global Ocean Color (GlobColour) dataset (Garnesson et al., 2019). GLORYS12 is a high-resolution reanalysis that provides daily and monthly data on SSS, SST, MLD, and SSH, among other variables. It is based on the NEMO platform and driven at the surface by ECMWF’s ERA-Interim and ERA5 reanalyses. Satellite altimetry, SST, sea ice concentration, and in situ temperature and salinity vertical profiles are assimilated using a reduced-order Kalman filter, with a 3D-VAR scheme that corrects large-scale biases (Lellouche et al., 2021).

2.4 [Model evaluation and prediction generation](#)

GlobColour is a combination of satellite observations of chl-a (alongside other ocean color variables) from multiple remote sensors. Chl-a concentrations are derived from reflectance using sensor-specific algorithm coefficients and then merged. Bias correction is applied directly to the chl-a field, improving consistency and reliability (Garneison et al., 2019; Maritorena et al., 2010). Since remotely sensed chl-a is limited by sunlight availability, we focus on the -60° to 80° latitudes. Figure 1 shows a brief overview of the methodology.

A summary of the methodology: Six months of time series spanning January 1998 to May 2017 are used for training twelve neural networks, each initialized on the starting month of the forecast.

2.5 Validation and prediction generation

We first validated our neural architecture approach by generating a set of 5-day predictions covering the 2017-2023-2018-2023 period using the reanalysis (GLORYS12) as input. Then, to generate the seasonal predictions We did this in order to assess performance at higher temporal resolution. However, to mimic the real-world application of the model, where the seasonal chl-a prediction would depend on the predicted physics as opposed to the reanalysis, we used 6-month forecasts from SEAS5, the seasonal forecasting system from the ECMWF (Johnson et al., 2019), which are provided at monthly resolution. Both approaches use a spatial resolution of $1/4^{\circ}$.

The forecast consists of a-an ensemble of 51 members, and it uses the IFS (Integrated Forecasting System) cycle 43r1 for the atmosphere, NEMO for the ocean, and LIM2 (Louvain-la-Neuve Sea Ice Model version 2) for sea ice. The system assimilates in-situ ocean observations (temperature, salinity), satellite altimetry, and snow cover using the OCEAN5 system. Initial conditions for the atmosphere and land surface come from ECMWF operational analyses, and forcings include CMIP5 historical and RCP 3-PD greenhouse gases (Johnson et al., 2019).

The SEAS5 (Johnson et al., 2019). The forecasts are available free of charge on the Copernicus Climate Data Store at a monthly, one-degree resolution (Copernicus Climate Change Service, 2018a, b). A high version of higher spatial resolution ($1/4$ degree) version $^{\circ}$) of the data was kindly provided by the ECMWF, but it is possible to obtain similar results by downscaling with a simple interpolation method.

The SEAS5 forecast has some minor differences with respect to the GLORYS12 reanalysis. To avoid these from biasing our evaluation of the model, we minimized them by calculating exhibits some biases relative to the reanalysis (see Fig. 2), which is expected given their different sources and purposes. To account for this, we computed the difference between the forecast and reanalysis climatologies, and then subtracting this difference from monthly climatologies of both datasets over the training period (1998-2017), then subtracted this mean bias from the SEAS5 data before using it as input-model input. This offset is computed once, and can be applied to all forecasts, including future ones.

We used the first 10 (out of 51) ensemble members from SEAS5 as input for the neural network ,and generated to generate six-month predictions spanning at a monthly resolution, using forecasts initialized from 01/2020-052020 to 12/2024. We then concatenated the n th month of each 6-month chl-a prediction to obtain five continuous time series \pm 2023. To evaluate performance across different lead times (i.e., the number of months after forecast initialization), we constructed six continuous time series by concatenating the corresponding month from each six-month forecast. Specifically, we extracted the first month

Mean bias between physical predictors, 2020-2023
(Seasonal forecast SEAS5, lead time 1) - (Reanalysis GLORYS12)

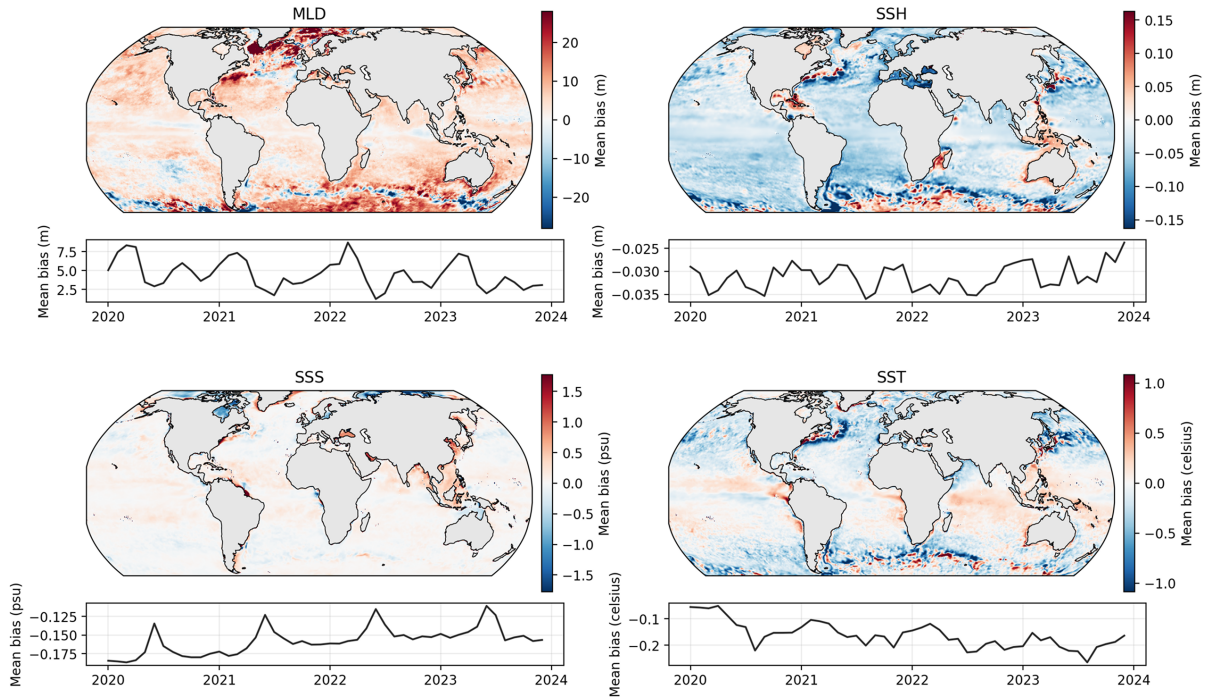


Figure 2. Mean biases between SEAS5 (monthly forecast, lead time 1) and GLORYS12 for the input variables over the 2020-2023 period. While correction is applied using historical climatological differences (see Methods), this figure is indicative of the systematic differences between the forecast and reanalysis.

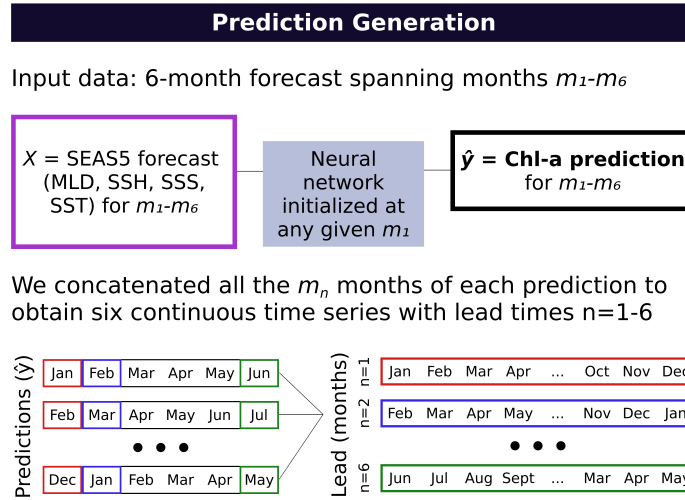
215 from each forecast to create a one-month lead time series (01/2020-12/2023 for one-month lead time, -), the second month
 from each forecast for a two-month lead time series (02/2020-01/2024 for lead-time two, etc), and so on. This is illustrated on
 Figure2 in Figure 3 for clarity. We compute SEAS5-based skill metrics over 01/2020-12/2023 (i.e., truncating any lead time
 series months that extend into 2024).

220 We evaluated performance by calculating the Anomaly Correlation Coefficient (ACC) for each longitude-latitude grid point
 as follows :-

$$ACC = \frac{\sum (r_i - c_r)(p_i - c_p)}{\sqrt{\sum (r_i - c_r)^2 \sum (p_i - c_p)^2}}$$

Where r_i and p_i are the reference (i.e. GlobColour) and predicted values, respectively, and c_r and c_p are the climatologies of the reference and predicted values, respectively.

A summary of the prediction-generation workflow: We concatenated the n th months from each prediction to generate continuous time



series with lead times 1-5.

Figure 3. A summary of the prediction generation workflow: We concatenated the n th months from each prediction to generate continuous time series with lead times 1-6.

We also report Root Mean Squared Error (RMSE), the Pearson correlation coefficient (r), normalized RMSE (NRMSE) and normalized Mean Absolute Error (NMAE). The latter two are calculated with respect to the standard deviation of the reference data:-

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum (r_i - p_i)^2}}{\sigma_r}$$

$$\text{NMAE} = \frac{1}{N} \sum \frac{|r_i - p_i|}{\sigma_r}$$

We analyzed performance by zooming into some open-ocean regions of interest (illustrated on Figure 3): the North Pacific (170°E-130°W, 25°N-45°N), Tropical Pacific (170°E-100°W, 10°S-10°N), South Pacific (170°W-100°W, 35°S-15°S), North Atlantic (70°W-20°W, 25°N-40°N), and Indian Ocean (55°E-95°E, 25°S-0°S), which are based on Park et al. (2019b), as well as. We also examined model skill in the Equatorial Atlantic (35°W-5°E, 5°S-5°N) and in selected large coastal ecosystems, specifically the California Current (25°N-45°N, 130°W-113°W), Humboldt Current (50°S-5°S, 85°W-69°W), South China Sea (1°N-25°N, 105°E-119°E), Gulf of Mexico (18°N-30°N, 100°W-80°W), Benguela Current (35°S-12°S, 5°E-20°E), Bay of Bengal (5°N-25°N, 80°E-98°E), Celtic-Biscay Shelf (43°N-55°N, 15°W-0°E), and Northeast Australian Shelf (25°S-10°S, 140°E-155°E). These regions are illustrated in Figure 4.

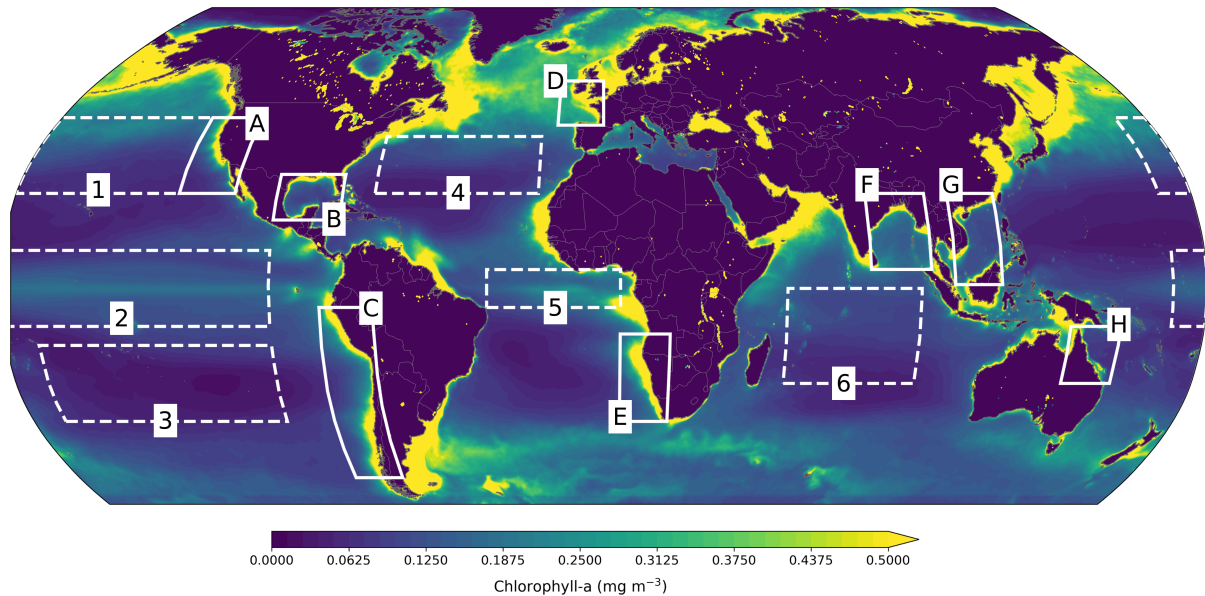


Figure 4. Regions where performance is evaluated. Open-ocean (dashed lines): North Pacific (A1), Tropical Pacific (B2), and South Pacific (C3), North (4) and Equatorial Atlantic (E5), and Indian Ocean (6). Coastal large ecosystems (solid lines): California Current (A), Gulf of Mexico (B), Humboldt Current (C), Celtic-Biscay Shelf (D), Benguela Current (E), and Equatorial Atlantic Bay of Bengal (F), South China Sea (G), and Northeast Australian Shelf (H).

We evaluated our results ~~with respect to the~~ against BIO4, a state-of-the-art global biogeochemical ~~analysis-BIO4 product~~ from Mercator Océan International (Copernicus Marine Service, 2023), ~~which~~ BIO4 is based on the PISCES model (Pelagic Interactions Scheme for Carbon and Ecosystem Studies), the biogeochemical component of the NEMO platform (Lamouroux et al., 2023; Aumont et al., 2015). ~~BIO4 models the~~ It simulates global cycles of carbon, oxygen, and key nutrients, and includes two functional groups ~~of both each of~~ phytoplankton and zooplankton, ~~generating and produces~~ a 3D global-scale representation of 24 representation of nearly 30 biogeochemical variables. ~~It provides both an analysis and a~~ BIO4 provides both analyses and 10-day forecast at depths of up to 5700m, ~~governed by internally consistent differential equations that capture the dynamics of ecosystems globally. BIO4 is forced depth. The model is driven by GLO12, a near real-time system and forecast based on the same physics reanalysis (near-real-time physical reanalysis based on GLORYS12) used for training, the training input physics~~ in our study. ~~Additionally, BIO4 assimilates chl-a observations from GlobColour, and it assimilates GlobColour observations~~ on a weekly basis (Lamouroux et al., 2024).

~~Since our focus is on~~ Our evaluation focused on the surface chl-a, ~~we evaluated our results against the surface output layer~~ of the BIO4 analysis. ~~Rather than a direct comparison, we use (rather than the forecast) due to differing forecast horizons.~~ We note that BIO4 as a benchmark, recognizing that it simulates a wide range of intereconnected biogeochemical processes across various depths, whereas our data-driven approach is specifically designed for surface biogeochemical variables and

processes beyond surface chl-a, while our neural network is specifically optimized for chl-a prediction and trained directly on observational data. We therefore use BIO4 as a contextual benchmark rather than a direct point of comparison.

255 We evaluate model skill primarily using RMSE and Pearson correlation computed on the raw (non-log-transformed) chl-a fields. All figures in this manuscript use non-log-transformed chl-a. Because different diagnostics emphasize different aspects of performance, we report two complementary RMSE definitions. For the regional time series shown in Figs. 11-14, we first spatially average each field over the region to obtain a region-mean time series, then compute RMSE between the ensemble-mean forecast and the GlobColour reference over 2020-2023 (with RMSE of the BIO4 analysis computed on the same dates). For the regional RMSE summaries (Figs. 10, 16-17), we compute RMSE within each region using all valid
260 grid-point samples over 2020-2023 either for a given calendar month (Figs. 16-17) or averaged across months (Fig. 10). For anomaly visualizations (Figs. 12, 14), anomalies are computed relative to a monthly historical climatology based on GlobColour (1998-2017). This provides a stable observation-based baseline while retaining systematic forecast biases.

3 Results

265 Chlorophyll-a concentration: Neural network (NN)-generated prediction with GLORYS12 (GL12) as input physics (green), and the reference observations (GlobColour) in black.

3.1 Model validation with reanalysis physics as input

To validate the neural ~~architecture~~ network architecture and isolate model performance from forecast uncertainties, we first generated a set of 5-day predictions covering the 2017-2023 period using the reanalysis (evaluated the model using GLORYS12) as input, illustrated in Figure 4. The spatially-averaged reconstructed time series has a RMSE of 0.01, NRMSE of 0.32,
270 NMAE of 0.24, and a correlation coefficient of ~~r~~ reanalysis as input physics. We generated predictions for 2018-2023 at a 5-day temporal resolution. Figure 5 shows that the globally averaged (spatial mean at each time step) time series of the neural network predictions (NN with Physics=0.95-G12) exhibits good visual agreement with the observations. The network captures both the amplitude and phasing of the dominant signal, validating its ability to learn the relationship between the four physical variables and biological productivity.

275 We then generated a 10-member ensemble of seasonal predictions using the

To assess spatial skill, Figure 6 presents a global map of Pearson correlation coefficients (r) computed at each grid point between the predictions and observations. Correlation values were computed for monthly-averaged data over the 2020-2023 period to facilitate comparison with the forecasting results, which have lower temporal resolution and only cover that temporal period. The correlation map shows strong spatial coherence, with positive correlation values across most of the global ocean.
280 Regions of particularly high correlation are found in the Northern oceans, while the Tropical and South Pacific show less skill.

3.2 Seasonal forecast performance

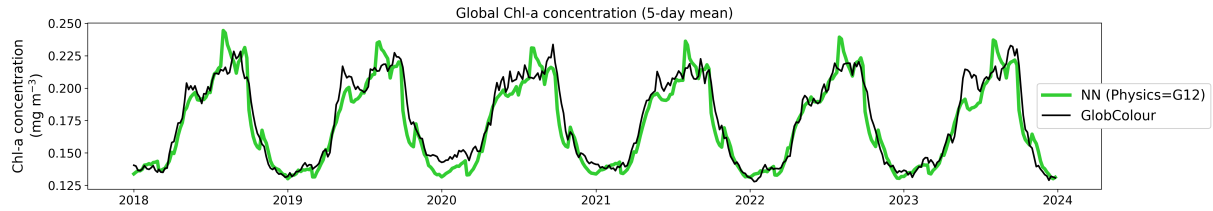


Figure 5. Globally averaged chl-a concentration time series from 2018 to 2023. The neural network predictions using GLORYS12 reanalysis as input physics (green line) are compared against GlobColour satellite observations (black line).

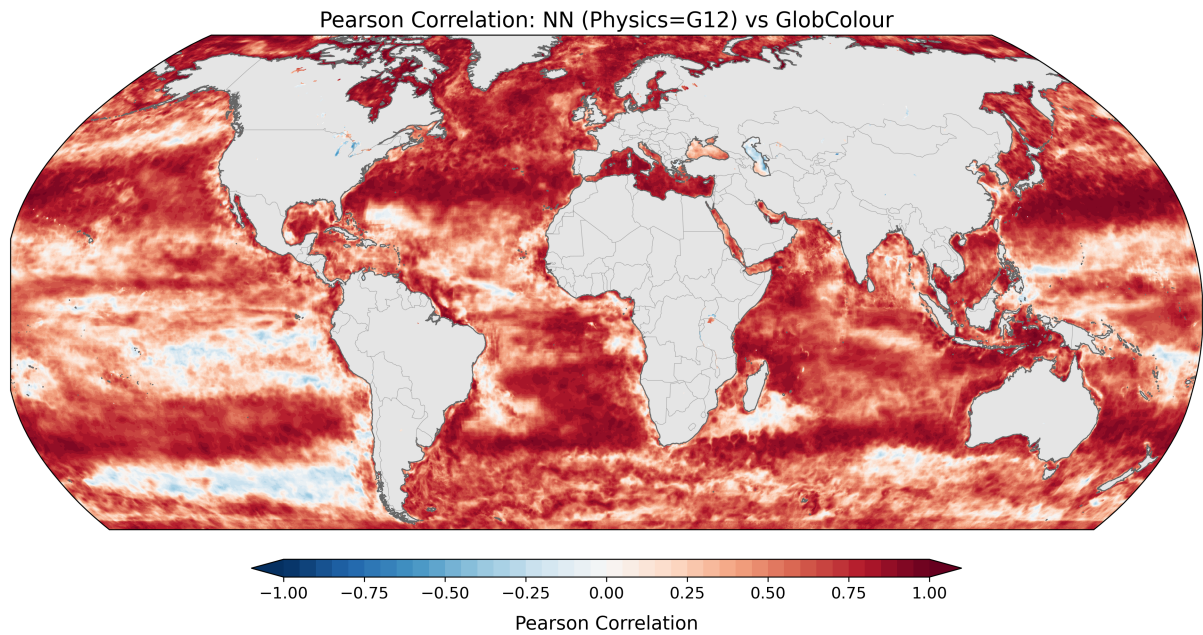


Figure 6. Spatial distribution of Pearson correlation coefficients between the neural network predictions (with GLORYS12 as input physics) and the observations. Correlation values are computed at each grid point for the monthly-averaged data over the 2020-2023 period for better comparison with forecasting results.

We next evaluated the model in a forecasting configuration by using SEAS5 seasonal forecasts as input. These six-month predictions span from Jan-Jun physics. All results presented below use the monthly model and the SEAS5 ensemble (10 of 2020-2023-51 members) as input. We note here that the chl-a reconstruction model is deterministic, and the ensemble spread arises solely from the SEAS5 physical ensemble.

We generated 6-month forecasts at monthly resolution for 2020-2023. Figure 7 displays the globally averaged time series for forecasts initialized in March, June, September and December, with individual ensemble member trajectories shown as thin lines and initializations indicated by colored markers. The ensemble forecasts are visually consistent with the observations, with individual member trajectories continuing to Dec-May of 2020-2024 (see Fig. 2 for details) at a 1/4 degree, monthly resolution. On Figure 5 we can see the ensemble of predictions initialized on January, April and June, with these months highlighted. These months correspond to lead-times one out of the six months of each forecast. While the predicted values tend to follow the observations closely, the model is not without its limitations. Figure 5 shows that while the SEAS5-driven neural network accurately captures the start and end of the bloom's global average, it tends to underestimate its peak. In contrast, the GLORYS12-driven predictions (Fig. 4) slightly overestimate the peak of the seasonal cycle, and shift the timing of the maxima in almost every year. observed chl-a concentrations throughout the 6-month forecast horizon, indicating little skill degradation with increasing lead time.

Seasonal chl-a forecasts (10 members) initialized on January, April and June generated using SEAS5 (S5) as input physics (shown cyan, magenta and yellow, respectively). The circular marks correspond to the lead-time one of each forecast.

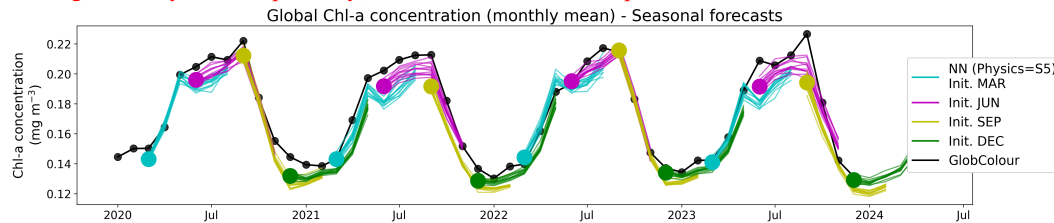


Figure 7. Globally averaged chl-a concentration time series from 2020 to 2023 reconstructed from the seasonal forecasts. The forecasts were initialized in March (cyan), June (magenta), September (yellow), and December (green), and the individual member reconstructions are shown as thin lines. Initialization (lead time 1) months are indicated by large colored markers. GlobColour observations are shown in black.

We concatenated all the n th-lead-time predictions to generate continuous time series (see Fig. 2). Figure 6 shows the Anomaly Correlation Coefficient (ACC) for lead-times one and six months. The ACC ranges from

Figure 8 illustrates spatial skill using the Pearson correlation coefficient between predictions and observations over the 2020-2023 period, shown for lead times 1 (indicating perfect correlation) to -1 (complete anticorrelation) with 0 signifying no correlation. The figure shows that the neural network-generated predictions have relatively high skill globally, with ACC values above 0.6 across various regions at lead-time one. As expected, and 6. Correlations remain spatially coherent but are reduced relative to the ACC decreases with lead-time due to GLORYS12 driven predictions (Figure 6). This reduction was expected from the use of forecasts as input, since they contain larger errors and introduce a distributional shift relative to the training

Pearson Correlation: NN (Physics=SEAS5, ens. mean) vs GlobColour

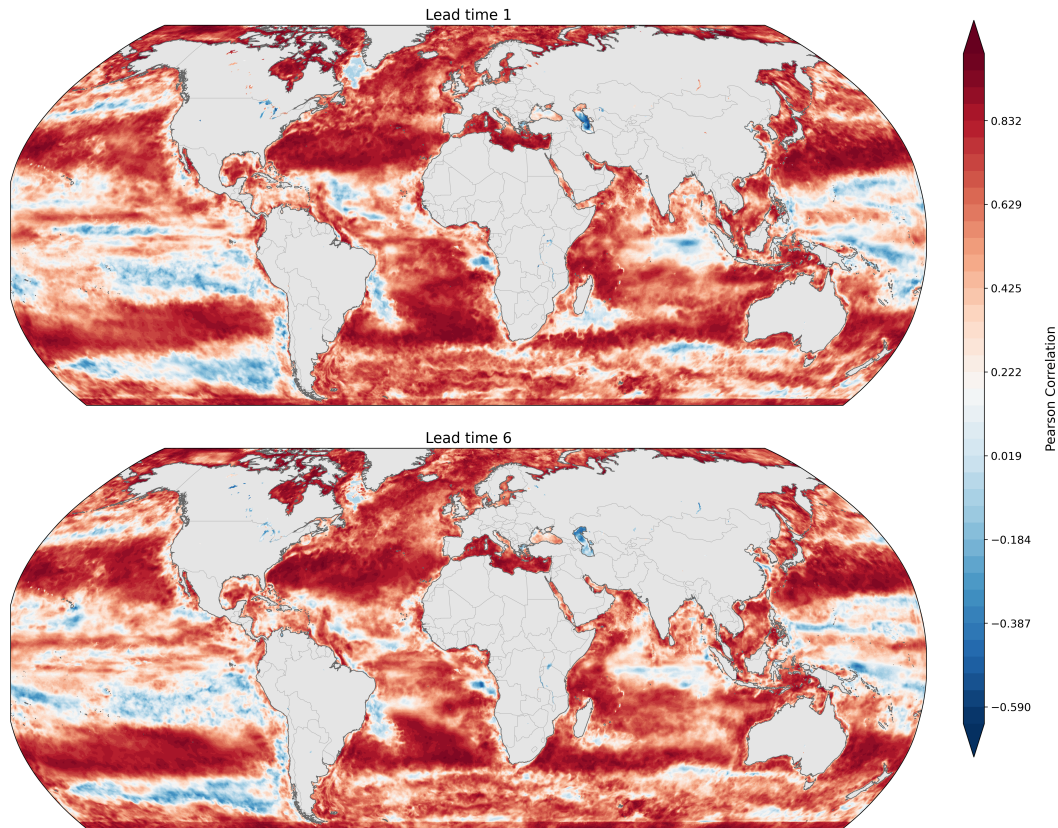


Figure 8. Spatial distribution of Pearson correlation coefficients for seasonal forecasts (SEAS5 used as input physics) at lead times 1 and 6 over 2020-2023. The ensemble-mean forecast is compared against GlobColour observations.

data (Figure 2). Notably, this reduction in skill persists in spite of the bias correction applied to the ~~increased uncertainty in the forecasted physics. However, despite the deterioration, positive correlations persist in numerous areas~~SEAS5 inputs (see Methods).

The largest degradations occur in the tropical and southern Pacific, the Indian Ocean, and plume-influenced regions near Brazil and to the south of the Gulf of Guinea. We also note that the southern Pacific (and, to a lesser extent, the tropical Pacific) exhibits relatively modest skill even in the GLORYS12 validation, suggesting that our particular modeling framework might be suboptimal for these regions, irrespective of the physics being used as input. Nonetheless, skill remains high in the subtropical gyre regions, where correlations are strong and spatially consistent. Additionally, the large-scale correlation patterns show little degradation from lead time 1 to lead time 6, indicating that the spatial skill of the predictions is relatively stable over the forecast horizon.

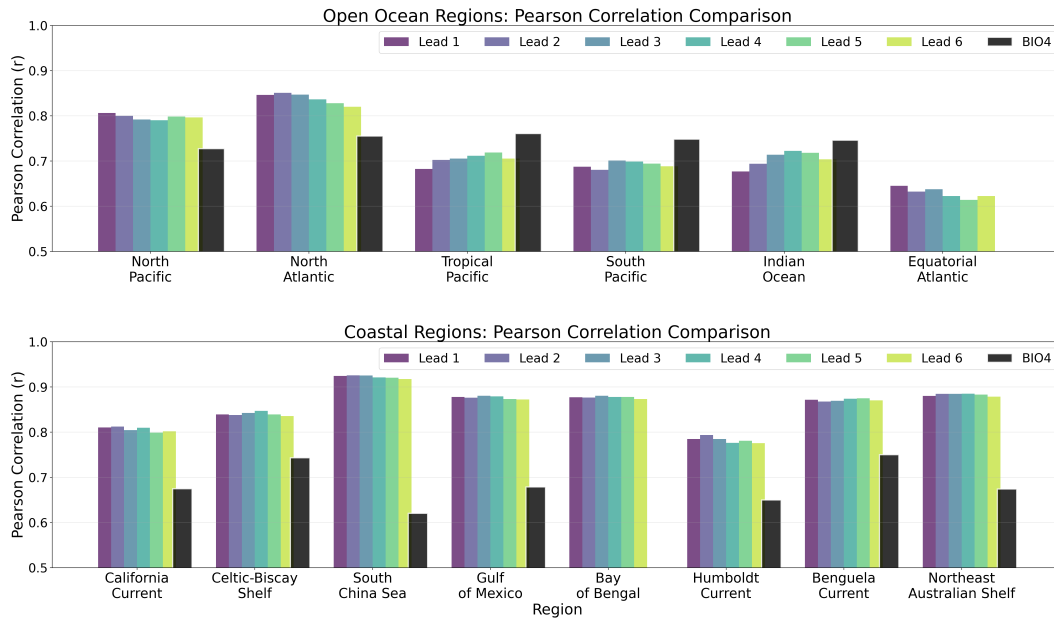


Figure 9. Anomaly-Comparative overview of Pearson correlation coefficient (ACC) coefficients for the generated predictions when using the SEAS5 (S5) forecast seasonal forecasts at lead-times one and six (months) lead times 1-6 over 2020-2023 for different regions. To better highlight predictive skill, Note that the color-scale limits were set to -0.8 and 0.8 y-axis is cut off at $r = 0.5$.

Monthly surface-level chl-a time series for each region. The reference dataset (GlobColour) is shown in black, the ensemble of the NN-generated predictions based on SEAS5 (lead-time one) in red, the prediction based on GLORYS12 in green, and the BIO4 output in blue.

Figure 7 shows the lead-time one predictions based on SEAS5 as input, those based on GLORYS12, the GlobColour observations, and the corresponding output from Figure 9 summarizes skill within each region from Fig. 4 using Pearson's r computed against the GlobColour reference. For each region (and lead time), r is computed using all valid grid-point pairs in the ensemble-mean forecast within the region for each calendar month (pooling 2020-2023), and then averaged across months. In the open-ocean regions (Fig. 9, top), correlations are generally higher than 0.7. The neural-network forecasts outperform BIO4 for the regions illustrated in the northern basins, consistent with the strong performance visible in Fig. 3. Figure 8 shows the Hovmöller diagrams for each region. These figures demonstrate that the neural network is able to capture the seasonal dynamics of the data across regions, regardless of the input physics. 8. They also outperform BIO4 in the Equatorial Atlantic, although skill there remains weak for the neural model too, as shown in Fig. 8. Performance is lower than BIO4 in the tropics and in the South Pacific and Indian Ocean, again consistent with the poor performance visible in Fig. 8.

In certain regions from Fig. 7, the transitions of the neural network's predictions can appear more abrupt than those of the observations and the numerical model. This is particularly noticeable in the Equatorial Atlantic, and the Tropical and South Pacific, where the time series have seemingly sharper increases and decreases in concentration. This abruptness may

335 be partially due to the way in which the time series is constructed, as the predictions are based on a concatenation of all the lead-one months (see Fig. 2). Since each prediction may contain its own biases, the process of stitching them together can introduce discontinuities, leading to these sudden shifts in concentration values. The magnitude of these errors can be further examined the coastal regions (Fig. 9, bottom), correlations are often higher (typically $r \geq 0.8$) and exceed BIO4 in all analysed regions, which is notable given the lack of explicit biogeochemical inputs and the challenges of coastal dynamics, although Fig. 8 shows that this performance is far from uniform. Across regions, correlations show limited sensitivity to lead time, with little degradation from lead 1 to lead 6.

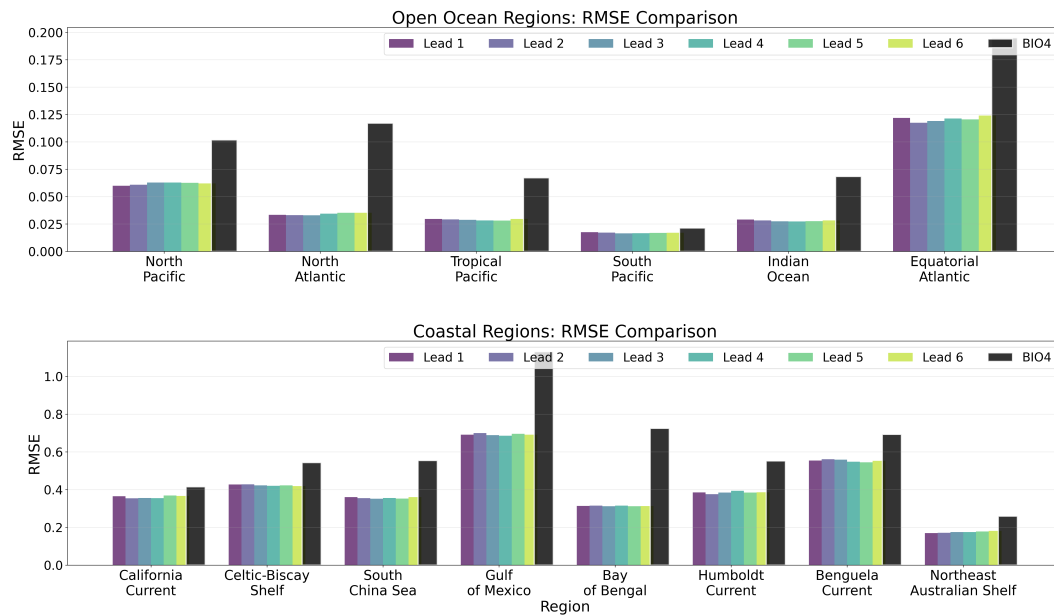


Figure 10. Comparative overview of RMSE for seasonal forecasts at lead times 1-6 over 2020-2023 for different regions, compared to the BIO4 analysis.

340 Figure 10 summarizes regional RMSE for the chl-a fields. For each region and lead time, similarly to r , the RMSE is computed over all valid grid cells of the ensemble-mean for each calendar month, and then averaged to obtain a single regional RMSE per lead time. The neural model exhibits lower RMSE than BIO4 across all regions considered, including regions where its correlation is lower (see Fig. 9). This indicates that, in those regions, the neural model is closer to the reference in absolute value even if it might be missing specific patterns (such as spatial shifts in mesoscale features), which can penalize correlation more strongly than RMSE. This is consistent with Fig. 15, where the neural predictions generally match the mean level and broad seasonal evolution of GlobColour but can show small timing/position offsets in bloom development. 345 Conversely, although correlations are often higher in coastal regions, RMSE is also higher on average (note the larger y-axis scale), consistent with the larger variability and dynamic range of coastal chl-a.

3.3 Further examining regional performance

350 Having assessed global-average skill, we now examine performance in the specific regions highlighted in Fig.8, which shows the chl-a values along the time (x-axis) and latitude (y-axis), and allows for a more detailed assessment of the evolution of the predictions spatially and temporally. 4 more in detail.

355 Figures 11 and 13 show regional time series for forecasts, this time initialized in January, April, July, and October. For each region, we compare the neural network predictions, the GlobColour reference observations, and the BIO4 analysis. Figures 12 and 14 show the corresponding anomalies, computed by subtracting a monthly GlobColour climatology (computed over 1998-2017) from each region to assess the ability to predict deviations from the seasonal cycle. We report the RMSE for each regional (and corresponding anomaly) time series, computed between the ensemble-mean forecast and the regional GlobColour time series over 2020-2023 (with BIO4 RMSE computed on the same dates). These time-series RMSE values quantify errors in the integrated regional signal after spatial averaging and thus complement the gridpoint-based RMSE reported in Fig. 10.

360 Figure 9 illustrates the evolution of RMSE, NRMSE, NMAE, and the correlation coefficient (r) with lead-time for each region. We also show

365 Across regions, the seasonal predictions generally reproduce the observed seasonal cycle and often track GlobColour more closely than BIO4 's performance (dotted lines) averaged over the period of study (01/2020-05/2024) for reference. The figure shows that in terms of magnitude and amplitude. Nonetheless, skill is not uniform and performance varies by region and initialization month, with larger discrepancies often occurring during the peak of the bloom. Note that the RMSE displayed in Figs 12 and 14 is identical to that of the neural network tends to have lower error than the corresponding raw-value figures (Figs 11 and 13) because the anomalies are computed by subtracting the same monthly GlobColour climatology from both the forecasts and the reference.

370 The anomaly plots reveal which deviations from the seasonal cycle the model successfully reproduces. The time series demonstrate that forecasts capture departures from the seasonal cycle with varying skill across initialization months and regions, and some cases exhibit consistent offsets or systematic biases. For example, in the South Pacific, predicted anomalies align more closely with BIO4 output in most regions. Additionally, than with observations, suggesting that shared errors may be tied to limitations in the physical input set, or may be predominantly driven by processes that we are either not modeling or struggling to represent.

375 Notably, in coastal regions such as the Celtic-Biscay Shelf, the South China Sea, the California Current, and the neural network's error rates and correlation coefficients remain relatively stable throughout the forecast period. This sustained performance with lead-time might seem contradictory to the decrease in ACC observed in Fig. 6, but it is likely that Bay of Bengal, the forecasts capture the overall trend and temporal evolution of the observed anomalies, despite not always matching absolute magnitudes.

380 Figure 15 shows Hovmöller diagrams, which visualize the evolution of the field as a function of time and one spatial dimension (latitude in this case). The plots summarize the spatiotemporal evolution of surface chl-a, making it easier to compare bloom timing, propagation, and amplitude across models. The SEAS5- and GLORYS12-driven neural-network predictions are

Open ocean seasonal forecasts
(Initialized in Jan, Apr, Jul and Oct)

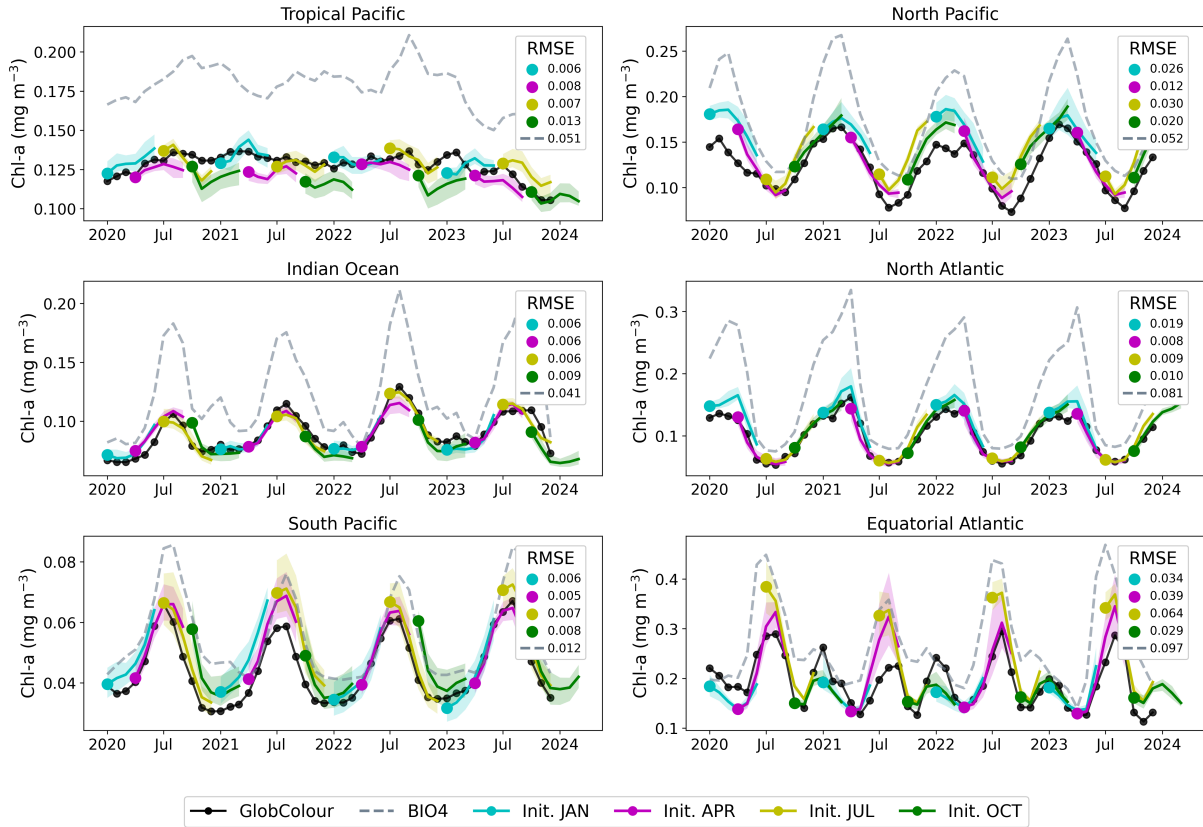


Figure 11. Hovmöller diagrams for each region: Regional, open-ocean forecasts. The solid black line represents the first column shows the average GlobColour reference dataset value for each region, while the second shows dashed lines indicate the corresponding BIO4 output. The solid colored lines represent the ensemble mean of average for the lead-time one SEAS5-based prediction neural network-predicted values initialized on January, the third shows the GLORYS12-based prediction April, July, and October, while the fourth shows shading represents the ensemble standard deviation. We report RMSE (computed after spatial averaging) for BIO4 output and each group of predictions.

Open ocean seasonal forecast anomalies
(Initialized in Jan, Apr, Jul and Oct)

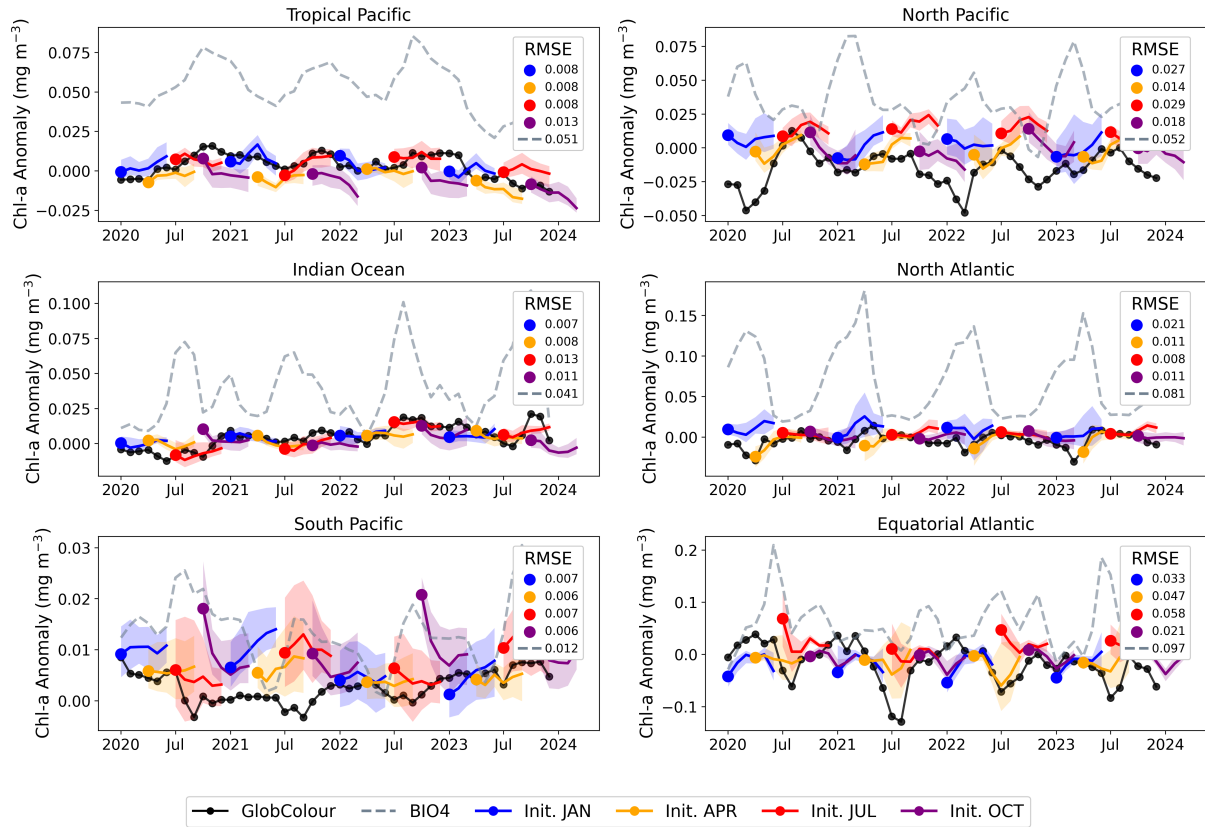


Figure 12. Metrics for different regions: RMSE Regional, NRMSE open-ocean forecasted anomalies. Anomalies are computed by subtracting the monthly GlobColour climatology (seasonal cycle, NMAE computed over 1998-2017) from each region-mean time series. The solid black line shows the GlobColour anomaly, dashed lines show the corresponding BIO4 anomaly, and correlation (r) solid colored lines show the ensemble-mean neural-network anomaly forecasts initialized in January, across different lead-times April, July, and October; shading indicates the ensemble standard deviation.

Coastal regions seasonal forecasts
(Initialized in Jan, Apr, Jul and Oct)

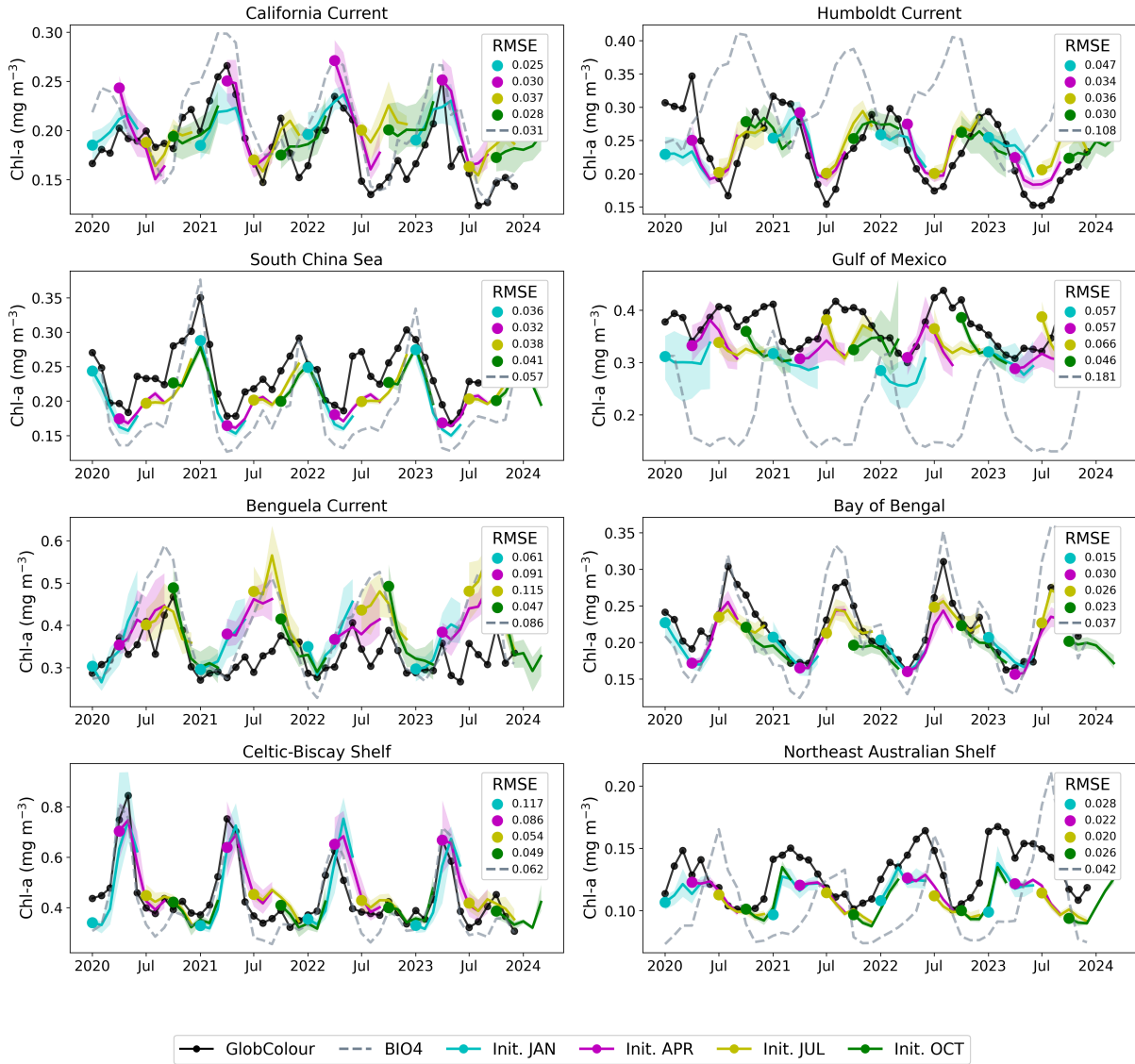


Figure 13. Regional, coastal forecasts. The figure shows GlobColour (black), BIO4 (dashed lines), and the ensemble average and standard deviation of the neural network predictions initialized on January, April, July, and October.

Coastal regions seasonal forecast anomalies
(Initialized in Jan, Apr, Jul and Oct)

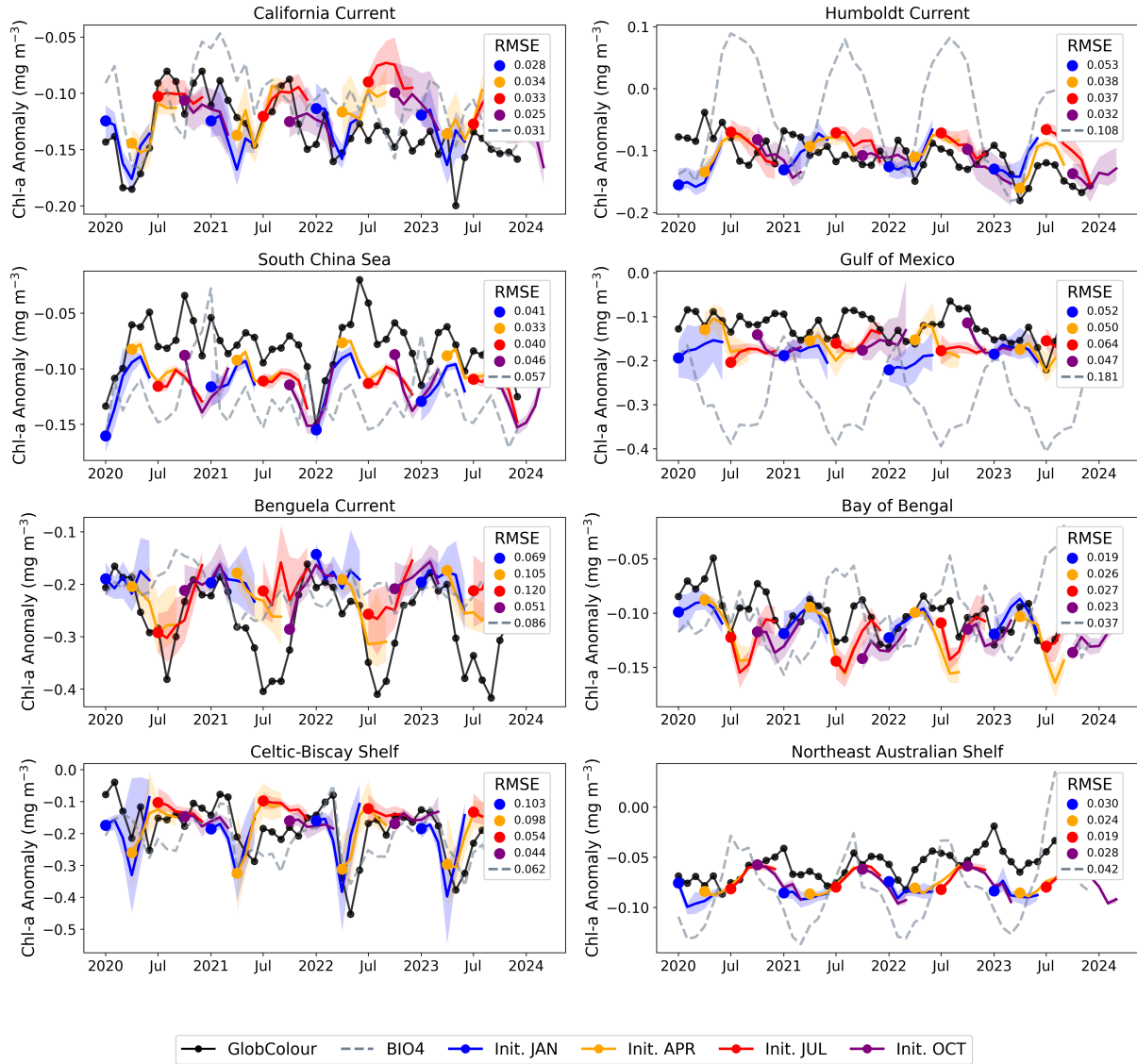


Figure 14. Regional, coastal forecasted anomalies. The figure shows the GlobColour anomaly (black), corresponding BIO4 anomaly (dashed lines), and the ensemble average and standard deviation of forecasted anomalies initialized in January, April, July, and October.

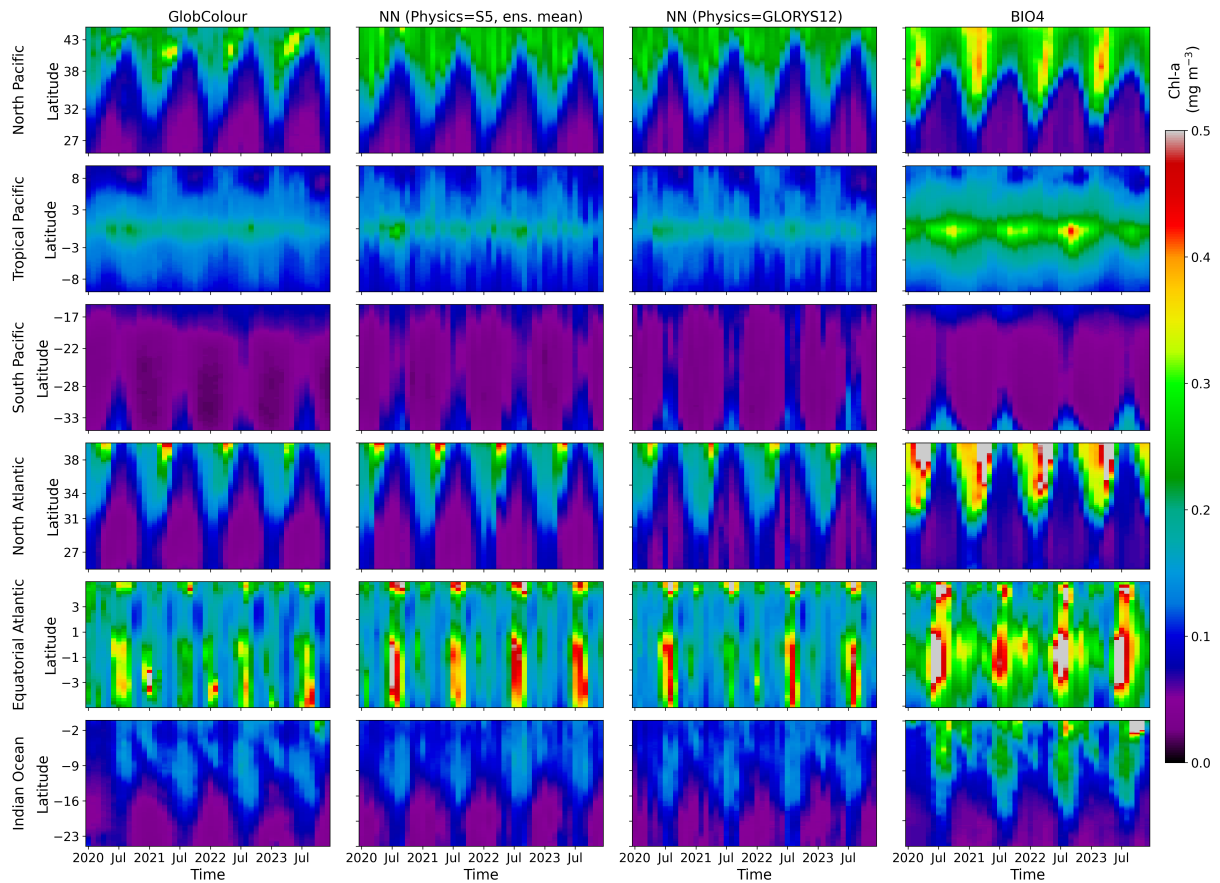


Figure 15. Hovmöller diagrams showing time vs. latitude for each open-ocean region. Columns show (1) the GlobColour reference, (2) the ensemble-mean lead-1 prediction forced by SEAS5 physics, (3) the corresponding prediction forced by GLORYS12 physics, and (4) the BIO4 output.

broadly similar across regions, but the neural-network’s ability to capture the strong seasonal dynamics in the data (Figs. 7 and 8) is compensating for Hovmöller format makes it possible to identify subtle shifts in the timing, latitude/extent, and amplitude of peak concentrations that are likely to contribute to the regional skill degradation highlighted in Fig. 8. BIO4’s tendency to overestimate chl-a magnitude, is apparent as systematically elevated values relative to GlobColour. These diagrams also highlight interannual variability: despite the dominant seasonal cycle, year-to-year differences in the timing and intensity of blooms are visible. The neural-network predictions exhibit variation across years rather than repeating a purely climatological seasonal pattern, even though these do not always align with the observed deviations from the seasonal cycle.

To investigate how error evolves through the seasonal cycle, Figures 16 and 17 show RMSE as a function of calendar month for each region. Here, RMSE is computed separately for each calendar month using all valid grid points within the region over 2020-2023 (i.e. not averaged across months like we do in Fig. 10).

Open Ocean Regions: Seasonal RMSE Comparison

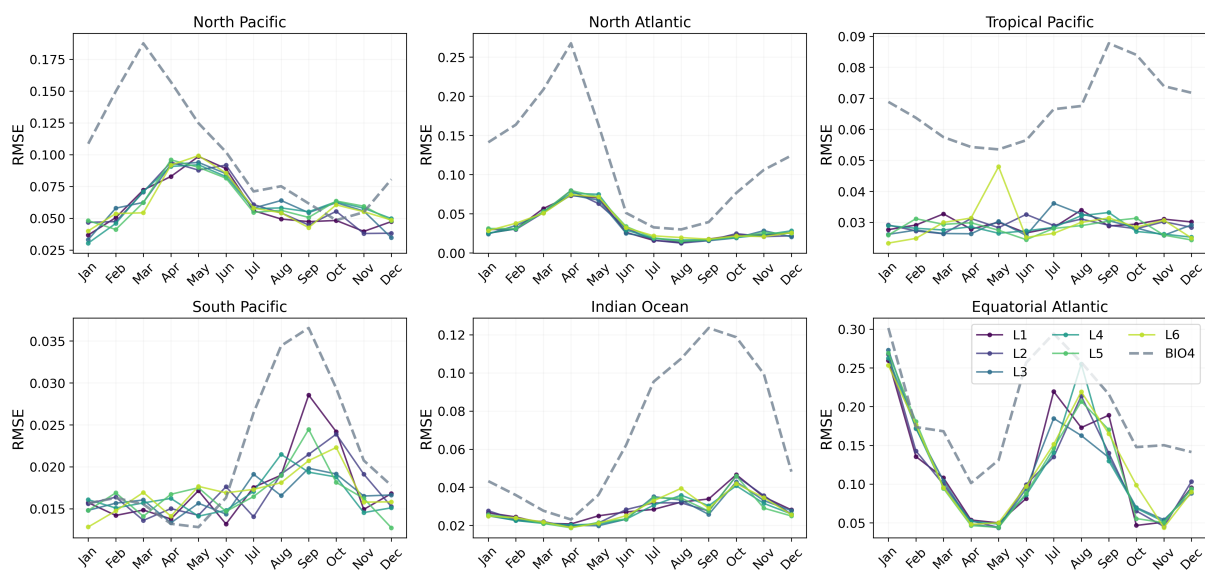


Figure 16. Seasonal evolution of the RMSE for each open-ocean region, comparing BIO4 (dashed lines) with the neural-network seasonal forecasts at lead times 1-6 (L1-L6).

Both the decrease in performance with respect to the anomalies. The decrease in ACC with increasing lead-time suggests that the neural network is less effective at predicting the exact magnitude and location of the anomalies over longer periods. However, Fig. 9 indicates that the model might still be good at capturing the general trends and overall patterns within each region, even as the forecast horizon increases and BIO4 exhibit increased errors during bloom peaks. Northern-hemisphere regions show larger RMSE in the first half of the year, while southern-hemisphere regions peak later, consistent with hemispheric seasonality. This behavior is expected because bloom peaks concentrate the largest amplitudes and strongest spatiotemporal gradients, in RMSE these timing offsets or amplitude damping near the peak translate into large squared errors. The fact that both the neural and process-based models show similar seasonality in error suggests that peak-bloom conditions may be intrinsically harder to predict, potentially reflecting sensitivity to rapidly evolving states and/or drivers that are not explicitly represented, rather than model class alone. However, this seasonal structure is not uniform across regions or between the two models. For example, in the Tropical Pacific the neural-network RMSE shows little seasonal variation compared to BIO4, while in the Humboldt Current BIO4 error has a secondary late-year peak that is not mirrored by the neural model, and the Benguela Current shows a different seasonal pattern. These differences suggest different sources of uncertainty and region-specific dynamics, and that accuracy might be constrained by different representations (or omission) of key drivers.

Coastal Regions: Seasonal RMSE Comparison

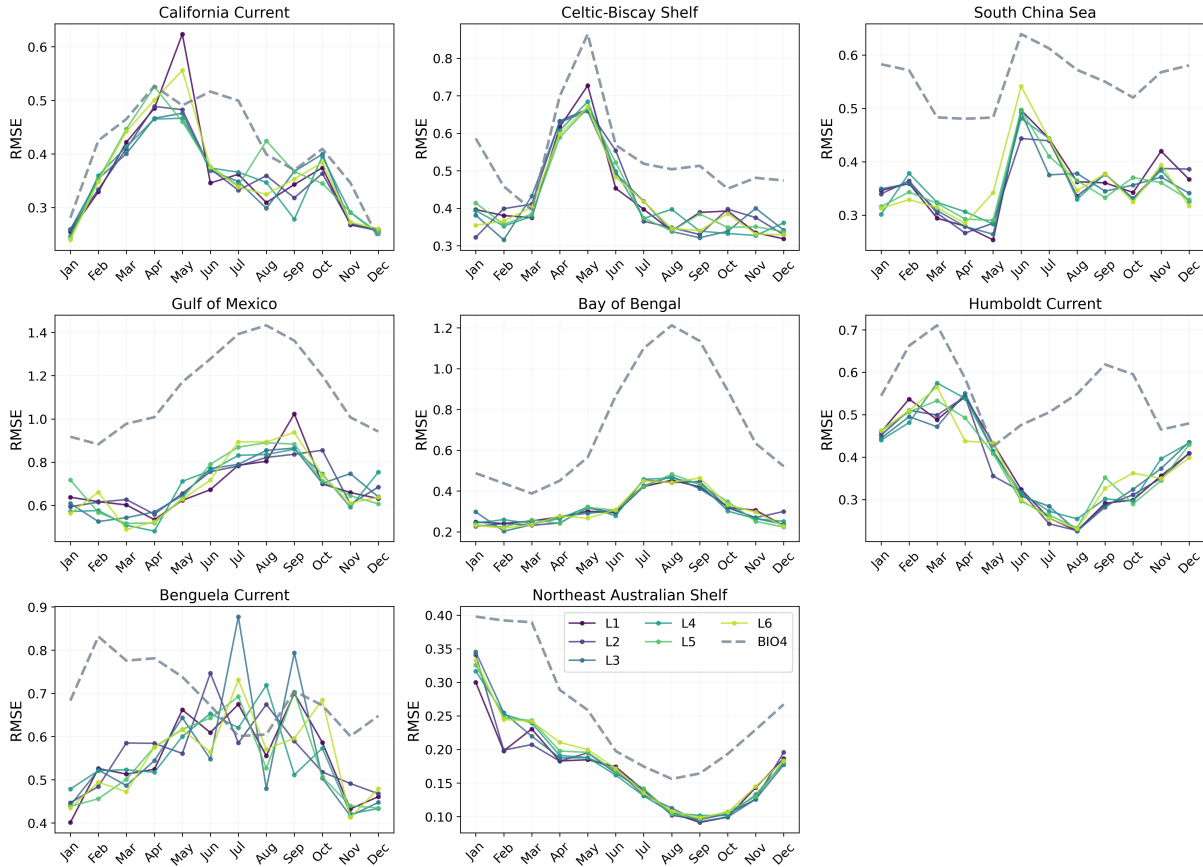


Figure 17. Seasonal evolution of the RMSE for each coastal region comparing BIO4 and seasonal forecasts at L1-L6

4 Discussion

We introduced a ~~simple neural network architecture that is capable of reconstructing surface,~~ lightweight, data-driven approach for reconstructing near-global surface chl-a from only four physical ~~variables predictors~~ (MLD, SSH, SSS and SST) ~~with relatively high accuracy and for extending this capability to seasonal forecasting by using forecasted physical ocean states~~ as input. Despite its simplicity, the method achieves high correlation and low error. ~~A key advantage of our approach is its computational efficiency: the model~~ across many regions, while remaining computationally efficient: the network can be trained in just few a few hours on a single GPU, and it generates 1/4-degree, 6-month predictions and can produce 1/4°, six-month forecasts within minutes. Our method maintains performance regardless of the source of the input physics data, and its accuracy can be further enhanced by correcting the inputs with a GLORYS12 climatology before use This makes the framework attractive for rapid generation of large-scale predictions, and in contexts where computational cost is a primary constraint.

Our approach aligns with previous studies that have demonstrated the potential of using physical ocean variables and machine learning for chl-a reconstruction (Sauzède et al., 2017; Martinez et al., 2020; Roussillon et al., 2023; Chen et al., 2024). By applying it to forecasted physical ocean states, we further expand the utility of this approach to forecasting, offering a simple yet useful framework for seasonal, near-global surface chl-a predictions. Our model focuses solely on surface chl-a, but this variable can be used as a proxy for phytoplankton biomass and variability in primary productivity, and it has the advantage of being more readily verifiable from observations. Skillful surface chl-a reconstructions and forecasts can help support the study of ecosystem responses to climate variability, and resource-management use cases that depend on these large-scale indicators. At the same time, we acknowledge that important subsurface processes and vertical structure are not represented in our approach, and so this framework would fall short whenever three-dimensional biogeochemical consistency is required.

~~We evaluated forecasting skill through the Anomaly Correlation Coefficient (ACC), error rates (RMSE, NRMSE, NMAE) and Pearson's correlation coefficient (r). While some degradation in ACC is observed with increasing lead time~~ Across the 2020-2023 evaluation period, the seasonal forecasts maintain spatially coherent skill from lead time 1 to lead time 6 (Fig. 8). Regionally, Pearson correlations generally remain high and show limited degradation with lead time (Fig.6), correlation and error rates on the raw time series remain satisfactory (9), while RMSE (Fig.9) (10) is consistently lower than BIO4, a state-of-the-art process based model. The relatively weak degradation that we see with increasing forecast horizon likely reflects that much of the skill is carried by large-scale, slowly varying components, and broad physical states that retain useful information across the 1-6 month range. The time-series and Hovmöller diagnostics (Figs. 11-13 and 15) indicate that the model reproduces the dominant seasonal cycle while exhibiting year-to-year variability, and that it captures the timing and magnitude of major bloom events in many regions. Errors nevertheless peak during bloom maxima (Figs.7 and 8 16 and 17), potentially compensating for the increased uncertainty in anomaly prediction over longer forecasting horizons due to heightened sensitivity to rapidly evolving physical conditions, or to drivers that are not explicitly represented in our input set.

440 Our approach retains useful skill in the large-scale evolution of surface chl-a, even when driven by forecasted physics that differ from the reanalysis used for training. The similarity between SEAS5- and GLORYS12-driven predictions (Fig. 15) suggests that the method is not tightly tied to a single source of physical data, although we do observe skill reduction relative to reanalysis-driven predictions in certain regions (Fig. 8). While this behavior is expected due to forecast errors (Fig. 2) potentially propagating into the chl-a reconstruction, a more detailed examination of these differences would be valuable in
445 future work. Such an analysis could identify where loss of skill is dominated by limits in input predictability versus model sensitivity, providing concrete targets for future model development. These could include region or regime-specific training strategies, domain adaptation approaches, or other input corrections aimed specifically at improving robustness.

4.1 Limitations

~~Despite its promising results, our approach has~~ Our framework is intentionally minimalist, and this design choice implies sev-
450 eral limitations. The decline in ACC with increasing lead-time suggests that while the model might effectively capture general seasonal patterns, it may struggle in ~~First, stricter diagnostics that emphasize interannual variability and small departures from the climatological seasonal cycle are highly phase-sensitive and can be penalizing of modest timing and spatial shifts in bloom features. This is expected because removing the seasonal cycle isolates a weaker, noisier signal, and because the monthly evaluation record over is relatively short (48 time steps). In addition, limited predictability in the forecast physical anomalies can further constrain skill. We therefore emphasize metrics (RMSE and Pearson correlation on the raw fields) that better reflect our main objective, which was reproducing realistic magnitudes and large-scale spatiotemporal evolution of surface chl-a in a resource-efficient manner. Performance in more complex scenarios where accurate reproduction of anomalies and outliers are crucial. This could also indicate that the neural network may be overfitting to certain historical patterns.~~ is crucial would likely require enhancements beyond the minimalist setup presented here. A more systematic treatment of interannual skill would also
455 require a more methodological assessment of SEAS5 skill with respect to anomalies, in order to evaluate how forecast errors might propagate into the chl-a reconstruction. A similar decomposition of skill for the process-based benchmark would enable a more principled comparison of performance between approaches.
460

~~When comparing our model to the BIO4 analysis, we find that BIO4 exhibits a more pronounced bias towards overestimation, a tendency that is also present in other traditional biogeochemical models (Rousseaux et al., 2021). Data-driven methods like ours, which learn directly from observations, might be better equipped to capture the actual pixel-value magnitudes of the data by avoiding the parameterizations that could be giving rise to these biases. However, the complexity of numerical models such as BIO4 also comes with key advantages. The numerical model accounts for a wide range of interconnected biogeochemical processes and it ensures, by design, internal consistency across all simulated variables. Second, we deliberately restrict predictors to four variables (MLD, SSH, SSS, and SST) that are available in both reanalyses and seasonal forecasts, and in order to keep the framework lightweight and readily deployable, but by doing so, we neglect drivers that can be crucial. These include irradiance (PAR), a key driver in light-limited regimes, winds and mixing variability not fully summarized by monthly MLD, and biogeochemical state variables (e.g. nutrients). As a result, it may be more reliable for extrapolating beyond previously observed conditions, whereas data-driven approaches can struggle with out-of-sample predictions if the training~~
465 When comparing our model to the BIO4 analysis, we find that BIO4 exhibits a more pronounced bias towards overestimation, a tendency that is also present in other traditional biogeochemical models (Rousseaux et al., 2021). Data-driven methods like ours, which learn directly from observations, might be better equipped to capture the actual pixel-value magnitudes of the data by avoiding the parameterizations that could be giving rise to these biases. However, the complexity of numerical models such as BIO4 also comes with key advantages. The numerical model accounts for a wide range of interconnected biogeochemical processes and it ensures, by design, internal consistency across all simulated variables. Second, we deliberately restrict predictors to four variables (MLD, SSH, SSS, and SST) that are available in both reanalyses and seasonal forecasts, and in order to keep the framework lightweight and readily deployable, but by doing so, we neglect drivers that can be crucial. These include irradiance (PAR), a key driver in light-limited regimes, winds and mixing variability not fully summarized by monthly MLD, and biogeochemical state variables (e.g. nutrients). As a result, it may be more reliable for extrapolating beyond previously observed conditions, whereas data-driven approaches can struggle with out-of-sample predictions if the training
470

475 ~~data lacks sufficient variability or if the model is not sufficiently robust, forecast errors will increase in regimes where chl-a~~
~~variability is strongly controlled by factors not captured in the input set, such as light limitation and nutrient supply. Several~~
~~studies suggest that interannual predictability in chl-a is closely linked to biogeochemical drivers, which can exhibit longer~~
~~predictability than physical variables alone (Séférian et al., 2014; Rousseaux et al., 2021; Park et al., 2019b; Ham et al., 2021; Frölicher et~~
~~. As with any modeling approach, trade-offs are inevitable. While biogeochemical variables can carry higher uncertainties, they~~
~~may also capture important feedbacks not fully represented by physical drivers. Future versions of this approach could therefore~~
480 ~~benefit from a more principled investigation of predictor uncertainty and sensitivity analyses to systematically incorporate any~~
~~additional relevant physical and biogeochemical inputs.~~

~~Our shallow neural architecture and the decision to limit the input to four variables were intended to create a lightweight~~
~~and efficient framework. However, Third, the neural network design is intentionally lightweight, but~~ this comes at the cost
of potentially failing to capture the full intricacy of the underlying processes, ~~and some applications may require a more~~
485 ~~sophisticated approach for greater robustness. This challenge of generalization can be further influenced by the use of the~~
~~MSE loss function in the training process. While~~. We used a simplified U-Net architecture and a month-conditioned strategy
implemented as twelve separately trained networks (one per initialization month) rather than a single unified model. While
this choice simplifies training across the seasonal cycle, it increases system complexity and can introduce month-to-month
inconsistencies because the mapping from physics to chl-a is learned independently for each network. A natural extension
490 would be to adopt a more robust architecture that explicitly conditions on month and lead time through a shared-backbone
design.

Furthermore, while MSE is a standard choice for regressive neural networks such as this one, its focus on minimizing error
can lead to overly smooth predictions that tend toward the mean. This can result in the loss of fine-scale detail in regions with
high variability, and an overall impact on the realism of the model's predictions.

495 ~~Finally, we present an "ensemble" of predictions that~~ Future work could mitigate this by adopting more robust learning
objectives and/or by increasing architectural capacity. In addition, the forecast "ensemble" presented here is derived from the
~~reconstruction of chl-a from the input forecast ensemble. By design, this approach inherits the uncertainties and biases of the~~
~~SEAS5 forecast, without accounting for the full range of uncertainty linked to the prediction of propagation of uncertainty from~~
~~the SEAS5 physical ensemble through a deterministic chl-a reconstruction. It does not explicitly represent chl-a uncertainties,~~
500 ~~and it inherits any shared biases and limits in the physical forecast.~~ A more ~~robust ensemble generation complete~~ approach
would require the inclusion of more comprehensive sources of uncertainty and the perturbation of these within the predictive
framework.

4.2 Future directions and improvements

~~The performance of our model is inherently linked to the quality and diversity of the training data. Research suggests that~~
505 ~~Finally, data-driven methods like ours, which learn directly from observations, can better reproduce observed chl-a has long-range~~
~~predictability (Rousseaux et al., 2021; Park et al., 2019b; Séférian et al., 2014; Ham et al., 2021; Frölicher et al., 2020), with interannual~~
~~skill closely linked to biogeochemical drivers, which may exhibit longer predictability than physical variables such as SST~~

(Séférian et al., 2014; Rousseaux et al., 2021; Park et al., 2019b). Similar to nutrient anomalies, SSS anomalies are influenced by MLD variability, and both of these have been identified as important factors in predictability (Fransner et al., 2020; Séférian et al., 2014;

510 =

This suggests that while our model may be capturing some of these key physical processes, it could be further improved by integrating biogeochemical variables as predictors. Expanding the training dataset to include a broader range of physical and biogeochemical drivers, as well as incorporating longer time series, vertical profile data, and more diverse sources of information, would almost certainly improve the model's predictive skill magnitudes and avoid some of the systematic biases that arise from uncertain parameterizations in process-based models (manifested as systematic overestimations relative to GlobColour in the case of BIO4). Nonetheless, this same minimalistic design, which emphasizes good skill in the raw fields in its current form, may also limit performance. Mechanistic biogeochemical models, by enforcing dynamical and biogeochemical consistency across variables, may be more robust for extrapolation beyond the observational regime, whereas purely data-driven approaches may degrade under distribution shift or in regimes that are poorly represented in the training data.

515

520

Another pathway for improvement could be scaling the neural architecture itself. This could be achieved by adding more of the same components (i.e. more convolutional layers, a deeper encoder-decoder structure, etc.), or by integrating it with other types of architectures. For example, diffusion-based approaches and probabilistic methods could enhance the ensemble generation process, while a more robust loss function or hybrid methods that leverage strengths of numerical techniques could increase performance and reliability.

525 5 Conclusions

Our resource-efficient, data-driven methodology demonstrates skillful-competitive results for the estimation of seasonal surface chl-a at a near-global scale. This is done by using MLD, SSH, SSS and SST data as input for a lightweight, U-Net-like neural network that is trainable in a matter of hours on a single GPU. Our method can be used to reconstruct historical chl-a patterns (i.e. by using a reanalysis as input), or to generate predictions based on easily accessible forecast data. ~~These predictions maintain~~ While future work is needed to better quantify and improve performance on interannual variability, these predictions show low error and high correlation to observations across different regions and ~~lead-times~~ lead times. The method is computationally efficient, and can be easily expanded to improve accuracy and extend its capabilities to other satellite-derived ocean-colour variables.

530

535

Author contributions. J.J, R.B and G.M.B conceived the study. G.M.B collected the data, developed the model, performed the data analysis and created the figures and original draft of the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. This work was ~~funded by the project DTO-Bioflow, which has received funding from the European Union~~partially supported by the European Union's Horizon Europe ~~Innovation Action under grant agreement No101112823.~~ Part of the work was supported ~~by the project NECCTON Programme under Grant Agreement No. 101112823 ("DTO-BioFlow) and by the NECCTON project ('New~~
540 Copernicus Capability for Trophic Ocean Networks"), ~~which has received funding from~~), funded under Horizon Europe RIA ~~under Grant~~
~~Number~~Grant No. 101081273. ~~This research was also supported by the French National Centre for Space Studies (CNES) under project~~
~~TOSCA SARGAT. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European~~
~~Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting~~
~~authority can be held responsible for them.~~

545 References

- Achterberg, E. P.: Grand challenges in marine biogeochemistry, *Frontiers in Marine Science*, 1, <https://www.frontiersin.org/articles/10.3389/fmars.2014.00007>, 2014.
- Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., and Hickey, J.: Machine Learning for Precipitation Nowcasting from Radar Images, <https://doi.org/10.48550/arXiv.1912.12132>, arXiv:1912.12132 [cs, stat], 2019.
- 550 Ansari, M. Y., Yang, Y., Balakrishnan, S., Abinahed, J., Al-Ansari, A., Warfa, M., Almokdad, O., Barah, A., Omer, A., Singh, A. V., Meher, P. K., Bhadra, J., Halabi, O., Azampour, M. F., Navab, N., Wendler, T., and Dakua, S. P.: A lightweight neural network with multiscale feature enhancement for liver CT segmentation, *Scientific Reports*, 12, 14 153, <https://doi.org/10.1038/s41598-022-16828-6>, 2022.
- Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, *Geoscientific Model Development*, 8, 2465–2513, <https://doi.org/10.5194/gmd-8-2465-2015>, publisher: Copernicus GmbH, 2015.
- 555 Barone, B., Coenen, A., Beckett, S., McGillicuddy, D., Weitz, J., and Karl, D.: The ecological and biogeochemical state of the North Pacific Subtropical Gyre is linked to sea surface height, *Journal of Marine Research*, 77, https://elischolar.library.yale.edu/journal_of_marine_research/478, 2019.
- Berardi, D.: 21st-century biogeochemical modeling: Challenges for Century-based models and where do we go from here?, *GCB Bioenergy*, 12, 774–788, <https://doi.org/10.1111/gcbb.12730>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcbb.12730>, 2020.
- 560 Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, *Biogeosciences*, 10, 6225–6245, <https://doi.org/10.5194/bg-10-6225-2013>, publisher: Copernicus GmbH, 2013.
- Browning, T. J. and Moore, C. M.: Global analysis of ocean phytoplankton nutrient limitation reveals high prevalence of co-limitation, *Nature Communications*, 14, 5014, <https://doi.org/10.1038/s41467-023-40774-0>, publisher: Nature Publishing Group, 2023.
- 565 Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., and Xue, K.: A machine learning approach to estimate chlorophyll-*a* from Landsat-8 measurements in inland lakes, *Remote Sensing of Environment*, 248, 111 974, <https://doi.org/10.1016/j.rse.2020.111974>, 2020.
- Chen, C., Chen, Q., Yao, S., He, M., Zhang, J., Li, G., and Lin, Y.: Combining physical-based model and machine learning to forecast chlorophyll-*a* concentration in freshwater lakes, *Science of The Total Environment*, 907, 168 097, <https://doi.org/10.1016/j.scitotenv.2023.168097>, 2024.
- 570 Chenillat, F., Illig, S., Jouanno, J., Awo, F. M., Alory, G., and Brehmer, P.: How do Climate Modes Shape the Chlorophyll-*a* Interannual Variability in the Tropical Atlantic?, *Geophysical Research Letters*, 48, e2021GL093 769, <https://doi.org/10.1029/2021GL093769>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL093769>, 2021.
- Copernicus Climate Change Service, C. D. S.: Seasonal forecast monthly averages of ocean variables, <https://doi.org/10.24381/cds.2f9be611>, 2018a.
- 575 Copernicus Climate Change Service, C. D. S.: Seasonal forecast monthly statistics on single levels, <https://doi.org/10.24381/cds.68dd14c3>, 2018b.
- Copernicus Marine Service, M. D. S.: Global Ocean Biogeochemistry Analysis and Forecast, <https://doi.org/10.48670/moi-00015>, 2023.
- Copernicus Marine Service, M. D. S.: Global Ocean Physics Reanalysis (GLORYS12), <https://doi.org/10.48670/moi-00021>, 2024.
- Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., and Garcia, R.: Incorporating Second-Order Functional Knowledge for Better Option Pricing, 2000.
- 580

- Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, *Nature Reviews Methods Primers*, 2, 1–21, <https://doi.org/10.1038/s43586-022-00154-2>, number: 1 Publisher: Nature Publishing Group, 2022.
- 585 Fernandez-Gonzalez, C., Tarran, G. A., Schuback, N., Woodward, E. M. S., Aristegui, J., and Maranon, E.: Phytoplankton responses to changing temperature and nutrient availability are consistent across the tropical and subtropical Atlantic, *Communications Biology*, 5, 1035, <https://doi.org/10.1038/s42003-022-03971-z>, 2022.
- Fransner, F., Counillon, F., Bethke, I., Tjiputra, J., Samuelsen, A., Nummelin, A., and Olsen, A.: Ocean Biogeochemical Predictions—Initialization and Limits of Predictability, *Frontiers in Marine Science*, 7, <https://doi.org/10.3389/fmars.2020.00386>, publisher: Frontiers, 2020.
- Frölicher, T. L., Ramseyer, L., Raible, C. C., Rodgers, K. B., and Dunne, J.: Potential predictability of marine ecosystem drivers, *Biogeosciences*, 17, 2061–2083, <https://doi.org/10.5194/bg-17-2061-2020>, publisher: Copernicus GmbH, 2020.
- 590 Garnesson, P., Mangin, A., Fanton d’Andon, O., Demaria, J., and Bretagnon, M.: The CMEMS GlobColour chlorophyll *a* product based on satellite observation: multi-sensor merging and flagging strategies, *Ocean Science*, 15, 819–830, <https://doi.org/10.5194/os-15-819-2019>, publisher: Copernicus GmbH, 2019.
- Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., and Simon, E.: Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, *Journal of Operational Oceanography*, 8, s168–s187, <https://doi.org/10.1080/1755876X.2015.1022350>, publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/1755876X.2015.1022350>, 2015.
- 595 Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J.-k., Chuprin, A., Ciavatta, S., Cipollini, P., Donlon, C., Franz, B., He, X., Hirata, T., Jackson, T., Kampel, M., Krasemann, H., Lavender, S., Pardo-Martinez, S., Mélin, F., Platt, T., Santoleri, R., Skakala, J., Schaeffer, B., Smith, M., Steinmetz, F., Valente, A., and Wang, M.: Satellite Ocean Colour: Current Status and Future Perspective, *Frontiers in Marine Science*, 6, <https://doi.org/10.3389/fmars.2019.00485>, publisher: Frontiers, 2019.
- 600 Ham, Y.-G., Joo, Y.-S., and Park, J.-Y.: Mechanism of skillful seasonal surface chlorophyll prediction over the southern Pacific using a global earth system model, *Climate Dynamics*, 56, 45–64, <https://doi.org/10.1007/s00382-020-05403-2>, 2021.
- 605 Hu, C., Feng, L., and Guan, Q.: A Machine Learning Approach to Estimate Surface Chlorophyll *a* Concentrations in Global Oceans From Satellite Measurements, *IEEE Transactions on Geoscience and Remote Sensing*, 59, 4590–4607, <https://doi.org/10.1109/TGRS.2020.3016473>, conference Name: IEEE Transactions on Geoscience and Remote Sensing, 2021.
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, *Geoscientific Model Development*, 12, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>, publisher: Copernicus GmbH, 2019.
- 610 Keiner, L. E.: Estimating oceanic chlorophyll concentrations with neural networks, *International Journal of Remote Sensing*, 20, 189–194, <https://doi.org/10.1080/014311699213695>, publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/014311699213695>, 1999.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, <https://doi.org/10.48550/arXiv.1412.6980>, arXiv:1412.6980 [cs], 2017.
- 615 Kolluru, S. and Tiwari, S. P.: Modeling ocean surface chlorophyll-*a* concentration from ocean color remote sensing reflectance in global waters using machine learning, *Science of The Total Environment*, 844, 157 191, <https://doi.org/10.1016/j.scitotenv.2022.157191>, 2022.
- Lamouroux, J., Perruche, C., Mignot, A., Paul, J., Szczypta, C., and Gutknecht, E.: Global Ocean Biogeochemistry Analysis and Forecast, Quality Information Document, <https://doi.org/10.48670/moi-00015>, 2023.

- Lamouroux, J., Perruche, C., Mignot, A., Paul, J., and Szczypa, C.: Global Production Center
 620 GLOBAL_ANALYSISFORECAST_BGC_001_028 Quality Information Document, Copernicus Marine Service,
<https://doi.org/10.48670/moi-00015>, 2024.
- Lellouche, J.-M., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L. G., Charly, R., Tony,
 C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D., and Pierre-Yves, L. T.: The Copernicus Global 1/12° Oceanic
 and Sea Ice GLORYS12 Reanalysis, *Frontiers in Earth Science*, 9, <https://doi.org/10.3389/feart.2021.698876>, publisher: Frontiers, 2021.
- 625 Lin, J., Miller, P. I., Jönsson, B. F., and Bedington, M.: Early Warning of Harmful Algal Bloom Risk Using Satellite Ocean Color and
 Lagrangian Particle Trajectories, *Frontiers in Marine Science*, 8, <https://doi.org/10.3389/fmars.2021.736262>, publisher: Frontiers, 2021.
- Ly, Q. V., Nguyen, X. C., Lê, N. C., Truong, T.-D., Hoang, T.-H. T., Park, T. J., Maqbool, T., Pyo, J., Cho, K. H., Lee, K.-S., and Hur, J.:
 Application of Machine Learning for eutrophication analysis and algal bloom prediction in an urban river: A 10-year study of the Han
 River, South Korea, *Science of The Total Environment*, 797, 149 040, <https://doi.org/10.1016/j.scitotenv.2021.149040>, 2021.
- 630 Maritorena, S., d'Andon, O. H. F., Mangin, A., and Siegel, D. A.: Merged satellite ocean color data products using a bio-optical model:
 Characteristics, benefits and issues, *Remote Sensing of Environment*, 114, 1791–1804, <https://doi.org/10.1016/j.rse.2010.04.002>, 2010.
- Martinez, E., Gorgues, T., Lengaigne, M., Fontana, C., Sauzède, R., Menkes, C., Uitz, J., Di Lorenzo, E., and Fablet, R.:
 Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach, *Frontiers in Marine Science*, 7,
<https://doi.org/10.3389/fmars.2020.00464>, publisher: Frontiers, 2020.
- 635 Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M.,
 Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L.,
 Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification
 over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, <https://doi.org/10.1038/nature04095>, 2005.
- Palacios, D. M., Hazen, E. L., Schroeder, I. D., and Bograd, S. J.: Modeling the temperature-nitrate relationship in the coastal upwelling
 640 domain of the California Current, *Journal of Geophysical Research: Oceans*, 118, 3223–3239, <https://doi.org/10.1002/jgrc.20216>,
<https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrc.20216>, 2013.
- Park, J., Kim, J.-H., Kim, H.-c., Kim, B.-K., Bae, D., Jo, Y.-H., Jo, N., and Lee, S. H.: Reconstruction of Ocean Color Data
 Using Machine Learning Techniques in Polar Regions: Focusing on Off Cape Hallett, Ross Sea, *Remote Sensing*, 11, 1366,
<https://doi.org/10.3390/rs11111366>, number: 11 Publisher: Multidisciplinary Digital Publishing Institute, 2019a.
- 645 Park, J.-Y., Stock, C. A., Dunne, J. P., Yang, X., and Rosati, A.: Seasonal to multiannual marine ecosystem prediction with a global Earth
 system model, *Science*, 365, 284–288, <https://doi.org/10.1126/science.aav6634>, publisher: American Association for the Advancement of
 Science, 2019b.
- Park, Y., Cho, K. H., Park, J., Cha, S. M., and Kim, J. H.: Development of early-warning protocol for predicting *chlorophyll-a* concen-
 tration using machine learning models in freshwater and estuarine reservoirs, Korea, *Science of The Total Environment*, 502, 31–41,
 650 <https://doi.org/10.1016/j.scitotenv.2014.09.005>, 2015.
- Polovina, J. J., Dunne, J. P., Woodworth, P. A., and Howell, E. A.: Projected expansion of the subtropical biome and contraction of the
 temperate and equatorial upwelling biomes in the North Pacific under global warming, *ICES Journal of Marine Science*, 68, 986–995,
<https://doi.org/10.1093/icesjms/fsq198>, 2011.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation,
 655 <https://doi.org/10.48550/arXiv.1505.04597>, arXiv:1505.04597 [cs], 2015.

- Rousseaux, C. S., Gregg, W. W., and Ott, L.: Assessing the Skills of a Seasonal Forecast of Chlorophyll in the Global Pelagic Oceans, *Remote Sensing*, 13, 1051, <https://doi.org/10.3390/rs13061051>, 2021.
- Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., and Martinez, E.: A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers, *Frontiers in Marine Science*, 10, <https://doi.org/10.3389/fmars.2023.1077623>, publisher: Frontiers, 2023.
- 660 Sadaippan, B., Balakrishnan, P., C.R., V., Vijayan, N. T., Subramanian, M., and Gauns, M. U.: Applications of Machine Learning in Chemical and Biological Oceanography, *ACS Omega*, 8, 15 831–15 853, <https://doi.org/10.1021/acsomega.2c06441>, publisher: American Chemical Society, 2023.
- Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., and Johnson, K. S.: 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, <https://doi.org/10.3389/fmars.2017.00128>, publisher: Frontiers, 2017.
- 665 Schneider, B., Bopp, L., Gehlen, M., Segsneider, J., Frölicher, T. L., Cadule, P., Friedlingstein, P., Doney, S. C., Behrenfeld, M. J., and Joos, F.: Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models, *Biogeosciences*, 5, 597–614, <https://doi.org/10.5194/bg-5-597-2008>, publisher: Copernicus GmbH, 2008.
- 670 Shah, K. D., Patel, D. K., Patel, H. A., and Nagrani, H. R.: EMED-UNet: An Efficient Multi-Encoder-Decoder Based UNet for Chest X-ray Segmentation, in: 2022 IEEE Region 10 Symposium (TENSYMP), pp. 1–6, <https://doi.org/10.1109/TENSYMP54529.2022.9864556>, iSSN: 2642-6102, 2022.
- Séférian, R., Bopp, L., Gehlen, M., Swingedouw, D., Mignot, J., Guilyardi, E., and Servonnat, J.: Multiyear predictability of tropical marine productivity, *Proceedings of the National Academy of Sciences*, 111, 11 646–11 651, <https://doi.org/10.1073/pnas.1315855111>, publisher: Proceedings of the National Academy of Sciences, 2014.
- 675 Sønnderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., and Kalchbrenner, N.: MetNet: A Neural Weather Model for Precipitation Forecasting, <https://doi.org/10.48550/arXiv.2003.12140>, arXiv:2003.12140 [physics, stat], 2020.
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M.: Learning Spatiotemporal Features with 3D Convolutional Networks, in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 4489–4497, IEEE, Santiago, Chile, ISBN 978-1-4673-8391-2, <https://doi.org/10.1109/ICCV.2015.510>, 2015.
- 680 Uz, B. M., Yoder, J. A., and Osychny, V.: Pumping of nutrients to ocean surface waters by the action of propagating planetary waves, *Nature*, 409, 597–600, <https://doi.org/10.1038/35054527>, publisher: Nature Publishing Group, 2001.
- Wang, H., Zhu, R., Zhang, J., Ni, L., Shen, H., and Xie, P.: A Novel and Convenient Method for Early Warning of Algal Cell Density by Chlorophyll Fluorescence Parameters and Its Application in a Highland Lake, *Frontiers in Plant Science*, 9, <https://doi.org/10.3389/fpls.2018.00869>, publisher: Frontiers, 2018.
- 685 Wenxiang, D., Caiyun, Z., Shaoping, S., and Xueding, L.: Optimization of deep learning model for coastal chlorophyll *a* dynamic forecast, *Ecological Modelling*, 467, 109 913, <https://doi.org/10.1016/j.ecolmodel.2022.109913>, 2022.
- Weyn, J. A., Durran, D. R., and Caruana, R.: Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data, *Journal of Advances in Modeling Earth Systems*, 11, 2680–2693, <https://doi.org/10.1029/2019MS001705>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001705>, 2019.
- 690 Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N.: Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models, *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002502, <https://doi.org/10.1029/2021MS002502>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002502>, 2021.

- Xu, D., Wang, T., Xing, X., and Bian, C.: The Relationship Between Nitrate and Potential Density in the Ocean South of 30°S, *Journal of Geophysical Research: Oceans*, 127, e2022JC018948, <https://doi.org/10.1029/2022JC018948>, 2022.
- 695
- Ying, C., Xiao, L., Xueliang, Z., Wenyang, S., and Chongxuan, X.: Marine chlorophyll-a prediction based on deep auto-encoded temporal convolutional network model, *Ocean Modelling*, 186, 102263, <https://doi.org/10.1016/j.ocemod.2023.102263>, 2023.
- Zhu, X., Guo, H., Huang, J. J., Tian, S., and Zhang, Z.: A hybrid decomposition and Machine learning model for forecasting Chlorophyll-a and total nitrogen concentration in coastal waters, *Journal of Hydrology*, 619, 129207, <https://doi.org/10.1016/j.jhydrol.2023.129207>,
700 2023.