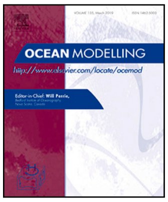
Ocean Modelling 201 (2026) 102707

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod



Two-phase CNN for model data fusion: Predicting 3D chlorophyll-a in the Mediterranean Sea

Teresa Tonelli [b,a,*], Gianpiero Cossarini [b], Luca Manzoni [b,a], Gloria Pietropolli [a,**]

[a] Department of Mathematics, Informatics and Geosciences, University of Trieste, Via Alfonso Valerio 12/1, 34127, Trieste, Italy
[b] National Institute of Oceanography and Applied Geophysics - OGS, Borgo Grotta Gigante 42/c, Sgonico, 34010, Trieste, Italy

ARTICLE INFO

Keywords:
3D convolutional neural network
Data fusion
Mediterranean Sea
Chlorophyll
BGC-argo float

ABSTRACT

Chlorophyll-a is a key indicator of marine ecosystem state and variability, monitored through satellite observations, in-situ measurements, and estimates from process-based models. Each source, however, suffers from intrinsic limitations such as incomplete coverage, multiple-source uncertainties, or simplified model representation. Reconstructing chlorophyll-a fields that accurately captures ecosystem variability, therefore, requires effective integration of these heterogeneous data through model–data fusion, which remains a major challenge. We introduce MuSt3Net, a deep-learning framework based on a Convolutional Neural Network (CNN) that performs model–data fusion through a sequential two-step learning strategy to predict the 3D distribution of chlorophyll-a in the Mediterranean Sea. MuSt3Net generates 3D chlorophyll-a fields by (i) learning the relationships between physical drivers and chlorophyll-a as produced by a process-based model, and (ii) integrating sparse in-situ BGC-Argo observations to propagate their information across the full 3D domain. This structured training scheme enables the network to preserve typical spatial patterns learned from the process-based model while integrating local corrections informed by observations. The reconstructed fields capture the characteristic spatial and seasonal variability of the Mediterranean Sea. Validation against independent BGC-Argo profiles confirms the model's skill, yielding a mean RMSE of $0.08\,\mathrm{mg\,m^{-3}}$. Winter surface blooms are reproduced with an error of approximately $0.1\,\mathrm{mg\,m^{-3}}$, and the deep chlorophyll-a maximum is represented with a vertical error below $10\,\mathrm{m}$. These results demonstrate the effectiveness of the proposed two-step fusion strategy for generating improved 3D biogeochemical reconstructions.

1. Introduction

Ocean biogeochemistry is a key component of the marine ecosystem, encompassing physical, chemical, and biological processes that regulate carbon and nutrient cycles. Among key biogeochemical variables, chlorophyll is particularly informative, as it reflects primary production, phytoplankton dynamics and their contribution to the biological carbon pump (Basterretxea et al., 2018; d'Ortenzio et al., 2014). Through its control on light absorption and short wave radiation, it also influences upper-ocean physics, thereby affecting heat redistribution and mixed-layer structure (Karagiorgos et al., 2025; Zhang et al., 2025). A comprehensive understanding of biogeochemical processes is essential for assessing the impact of climate change on the oceans and for supporting sustainable management of marine resources. However, accurately describing and predicting biogeochemical processes remains challenging due to (*i*) the complexity and interdependence of the factors that drive them and (*ii*) the limited number of observations capturing the spatiotemporal variability of biogeochemical variables (Russell et al., 2014).

Numerical ocean models, traditionally used for their ability to simulate the coupled physical–biogeochemical system, represent these processes through a hierarchy of parameterizations and governing equations (Kantha and Clayson, 2000). Although they provide indispensable tools for prediction and process understanding, the high dimensionality and complexity of biogeochemical interactions often require simplifying assumptions that can degrade model accuracy, particularly in regions characterized by strong mesoscale and submesoscale variability (Fennel et al., 2022).

To reduce the computational cost of running fully coupled simulations, surrogate models have emerged as efficient alternatives that emulate the behavior of more complex numerical systems (Asher et al.,

---

[*] Corresponding author at: National Institute of Oceanography and Applied Geophysics - OGS, Borgo Grotta Gigante 42/c, Sgonico, 34010, Trieste, Italy.
[**] Corresponding author.

E-mail addresses: ttonelli@ogs.it (T. Tonelli), gcossarini@ogs.it (G. Cossarini), lmanzoni@units.it (L. Manzoni), gloria.pietropolli@units.it (G. Pietropolli).

https://doi.org/10.1016/j.ocemod.2026.102707
Received 20 May 2025; Received in revised form 11 February 2026; Accepted 16 February 2026

2015; Bocquet, 2023). These models capture key features of high-dimensional numerical models while significantly reducing computational demand. Surrogates are typically built using data-driven function approximation techniques (Kleijnen, 2009), or lower-fidelity models that retain the main processes of the original system with reduced complexity (Forrester et al., 2007; Razavi et al., 2012).

Observations also play a fundamental role, as they are essential for model evaluation, parameter tuning, and skill assessment (Fennel et al., 2022; Stow et al., 2009). Among the available sources of ocean observations, BGC-Argo floats have become particularly important due to the large amount of physical and biogeochemical profiles they provide, improving both temporal coverage and insight into ocean interior dynamics (Miloslavich et al., 2019). However, their 1D nature and high operational cost can limit their spatial coverage, preventing comprehensive 3D domain representation.

The complementarity between models and observations makes their integration a central element of operational oceanography (Carrassi et al., 2018; Schiller et al., 2018). In this context, data fusion refers to the combination of heterogeneous datasets into coherent spatial and temporal fields (Castanedo, 2013; Hall and Llinas, 1997; White et al., 1991), and data assimilation (DA) represents a specific form of data fusion in which model outputs are adjusted using observations to improve estimates (Brajard et al., 2021; Carrassi et al., 2018; Teruzzi et al., 2014). Due to challenges such as model complexity, high dimensionality, approximations included in DA scheme, and the computational cost of repeated simulations (Hoteit et al., 2018; Howard et al., 2024), there is growing interest in refining existing schemes and exploring alternative data-fusion strategies (Gettelman et al., 2022; Ma et al., 2023).

In this context, Machine Learning (ML) and Deep Learning (DL) have become increasingly relevant (Gao et al., 2020; Shrestha and Mahmood, 2019; Tonelli et al., 2025; Q. Zhao et al., 2024). Neural networks (NN), the core architecture behind most DL models, have been widely applied to a range of oceanographic tasks, including ocean turbulence (Zanna and Bolton, 2021), predicting surface temperature (Kartal and Sekertekin, 2022), increasing model resolution (Ducournau and Fablet, 2016), and developing model data fusion algorithms (Brajard et al., 2021; Howard et al., 2024). Recent studies further highlight the growing potential of ML and DL for reconstructing three-dimensional (3D) and four-dimensional (4D) chlorophyll fields. For instance, Gaussian-activation deep neural networks have been used to resolve upper-ocean 3D chlorophyll structure at high spatial resolution (X. Zhao et al., 2024), and machine-learning approaches have been developed to estimate 3D chlorophyll distributions in the South China Sea (Zheng et al., 2024). Hybrid CNN–LSTM architectures have also been proposed to jointly predict physical and biogeochemical 4D fields in the Mediterranean Sea (Sammartino et al., 2025).

In this work, we present a DL framework based on a convolutional neural network (CNN) that integrates surrogate modeling and data fusion within a single architecture to predict the 3D distribution of chlorophyll from physical 3D fields and sparse observations. Our model combines deterministic biogeochemical model outputs with in situ BGC-Argo observations through a two-phase training strategy. In contrast to existing approaches (Sammartino et al., 2025), BGC-Argo profiles are integrated only in the second phase, allowing the network to first emulate the biogeochemical model, learning stable and physically consistent dynamics, i.e. key features of mesoscales, and then refine the reconstruction using observational information. This sequential design also enables flexible retraining, as new float profiles can be integrated without reinitializing the model, and additional data types can be integrated through the inclusion of an extra sequential training phase.

The Mediterranean Sea serves as an ideal testbed due to the availability of validated physical and biogeochemical reanalyses (Cossarini et al., 2021; Escudier et al., 2021), a higher-than-global-average coverage of BGC-Argo profiles (Amadio et al., 2024; Cossarini et al., 2019), and the pronounced variability in chlorophyll distributions, both horizontally (Lazzari et al., 2012; Volpe et al., 2018) and vertically (Lavigne et al., 2015).

This paper is organized as follows. Section 2 introduces the model and observation datasets, Section 3 describes the NN architecture and two-phases training procedure, and Section 4 outlines the experimental settings. Results (Section 5) and Discussion (Section 6) demonstrate the feasibility of our novel method and final thoughts are provided in Section 7.

2. Datasets

Our approach is implemented in the Mediterranean Sea, a semi-enclosed basin with a rich collection of numerical models, satellite observations, and in-situ data (Cossarini et al., 2019; Salon et al., 2019). Specifically, we use a validated numerical model (Section 2.1) and in-situ BGC-Argo float data (Section 2.2).

2.1. Numerical model data

Numerical models are the traditional approach for representing the 3D distribution of biogeochemical variables (Fennel et al., 2022; Gregg, 2008; Gutknecht et al., 2022). This study uses a coupled physical-biogeochemical (BGC) numerical model designed to simulate ocean biogeochemistry and its relationship with physical dynamics.

Specifically, we employ the output of the MedBFM model (Cossarini et al., 2021; Lazzari et al., 2016; Salon et al., 2019), which is the reference model for the Copernicus Marine Service of the Mediterranean Sea (Coppini et al., 2023). MedBFM is a numerical system model developed by OGS, which integrates the transport component (OGSTM) and the biogeochemical dynamics (BFM) and also includes a data assimilation scheme (Amadio et al., 2024; Cossarini et al., 2019; Teruzzi et al., 2014). The BFM component describes the biogeochemical cycles of carbon, nitrogen, phosphorus, and silicon across different ecosystem compartments, including dissolved inorganic, living organic, and non-living organic matter (Lazzari et al., 2012).

The model output used in the proposed model-data fusion architecture consists of a hindcast simulation of the period 2019–2021, whose setup (forcing, boundary, initial condition, model configuration) is described in Coppini et al. (2023) with the exception that data assimilation has not been included. Additionally, the physical 3D variables (i.e., velocity, temperature, salinity, sea level height) are provided by the MedFS model (Coppini et al., 2023), which features NEMO ocean dynamics, WWIII wave component, and OceanVar data assimilation of sea level anomaly and temperature and salinity profiles. The MedBFM hindcast simulation is forced by the MedFS output, which ensures consistency between physical and biogeochemical spatio-temporal dynamics.

Data assimilation was not applied in MedBFM so that biogeochemical observations are incorporated only within our data fusion step. Conversely, MedFS was used with data assimilation to provide the most accurate physical fields.

2.2. BGC-Argo float observations

The observations used in the proposed model-data fusion architecture come from the BGC-Argo program (Bittig et al., 2022, 2019), specifically from Argo floats collecting biogeochemical variables (BGC-Argo floats).

We used BGC-Argo S-profile data for the Mediterranean Sea, downloaded from the Coriolis Argo GDAC ((Argo, 2000), last accessed in July 2022) and we consider only Delayed Mode (DM) and Adjusted Real-Time Mode (RT) data for the period January 1, 2019 to December 31, 2021, ensuring a larger proportion of high-quality DM data.

The quality-checked BGC-Argo dataset used in this study is publicly accessible at Amadio et al. (2023). The dataset includes only complete

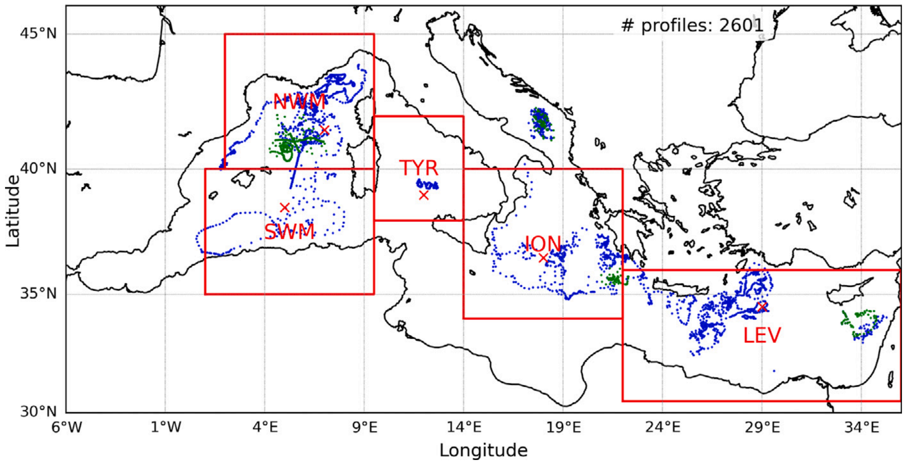


Fig. 1. Map of the Mediterranean Sea with the location of the used BGC-Argo floats: blue points refer to floats used during the train–test, while the green one represents the one used for the external validation procedure. The red blocks identify the different sub-basins, and red crosses represent the location centers used to estimate the temporal evolution of chlorophyll field per region.

profiles with Quality Flags 1 (good data), 2 (probably good data), 5 (value changed), and 8 (interpolated) for temperature, salinity, and chlorophyll. Profiles with a vertical resolution exceeding 10 m were averaged to a uniform 10-meter resolution. To further smooth out small fluctuations, we applied a moving weighted average using the two upper and two lower neighboring points, with weights determined by a Gaussian function of the distance to the central point.

The dataset includes floats distributed across the entire Mediterranean basin and covering different seasons. In the period January 1, 2019 to December 31, 2021, 2826 float profiles from 26 float devices are registered. The spatial distribution of the floats after quality control is reported in Fig. 1, which shows a higher float's presence in the NorthWest (NWM), Ionian (ION) and Levantine (LEV) Seas, while remaining poor in the SouthWest (SWM) and the Tyrrhenian (TYR) Sea.

3. MuSt3Net: model data fusion for 3D chlorophyll integration

This section presents MuSt3Net (Multiple-Steps 3D Convolutional Neural Network), a two-stage learning framework based on 3D CNNs for reconstructing three-dimensional concentration fields in marine environments. The model takes as input a concatenation of 3D tensors describing physical oceanographic variables together with sparse observations of the target variable, and produces a fully reconstructed 3D field. Although demonstrated here for chlorophyll, the framework is directly applicable to any ocean variable sampled by autonomous observing systems.

The methodology follows a structured approach, with input preprocessing detailed in Section 3.1, the specifics of the architecture described in Section 3.2, and the loss function outlined in Section 3.5. The core of this framework lies in its two-phase training procedure (Section 3.3). In the first phase, MuSt3Net learns spatial–temporal structure from the numerical model fields; in the second, it incorporates real float observations to refine and adapt the reconstruction. This sequential design leverages the full-domain MedBFM fields as a strong prior (Cossarini et al., 2019), allowing the network to capture large-scale patterns before being refined by sparse but highly informative in situ profiles.

3.1. Input preprocessing

The training dataset for MuSt3Net consists of discretized oceanographic fields derived from a numerical model, represented as 3D tensors on a fixed horizontal–vertical grid, along with sparse observations of the target variable discretized on the same grid. Each grid cell corresponds to a spatial region, analogous to pixels in image processing, while tensor channels store ocean variables instead of RGB values (Pietropolli et al., 2022).

All input and output variables are standardized by subtracting the mean and dividing by the standard deviation, which stabilizes the optimization and accelerates convergence. This normalization standardizes the feature scales, preventing certain variables from dominating the optimization process (Hastie et al., 2009; LeCun et al., 2002). We did not apply a log-transform to chlorophyll, as normalization was sufficient for stable training.

Given the sparsity of observation data, missing values are replaced with the spatial mean computed over the available values at the time step, ensuring numerical robustness while preserving the influence of real observations (LeCun et al., 2002). A 3D land–sea mask is applied to all inputs to prevent land points from affecting the reconstruction. This binary mask (1 for ocean, 0 for land) restricts loss computation to relevant oceanic regions (Wang et al., 2024), ensuring that the network focuses exclusively on marine areas and avoids artificial error minimization over land pixels.

3.2. MuSt3Net architecture

The 3D CNN used in MuSt3Net architecture is adapted from the model proposed in Roussillon et al. (2023), originally designed for chlorophyll reconstruction through its interactions with physical oceanographic variables. We refine this baseline through hyperparameter tuning. The same architecture is employed in both training phases. In the second phase, the network is *fine-tuned*: rather than training a new model from scratch, we continue training the weights obtained in the first phase so that the representation learned from model-based fields is adjusted to match the statistical characteristics of in-situ profiles (Cetinic et al., 2018; Kumar et al., 2022; Radenović et al., 2018).

The CNN architecture comprises six 3D convolutional layers, and its details are reported in Table 1. Each convolution is followed by batch normalization (Santurkar et al., 2018; Bjorck et al., 2018), which stabilizes training by enforcing zero-mean, and by an activation function. To improve generalization, dropout layers (Baldi and Sadowski, 2013; Hinton et al., 2012) are applied after each convolutional block. The final layer outputs a single 3D field representing the predicted chlorophyll distribution.

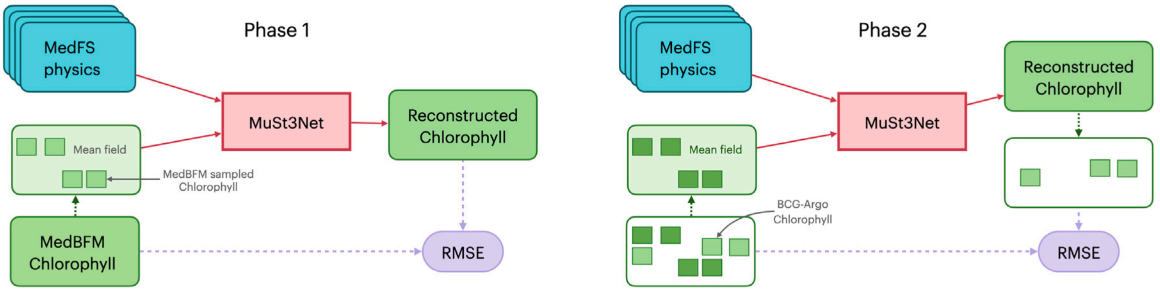


Fig. 2. Structure of the proposed two-phase training procedure showing inputs (blue and light green box), output (green box), and loss function (violet box).

Table 1
Architecture of MuSt3Net. BN refers to Batch Normalization, while the dropout probability d_r is set to 0.15, as reported in Table 2.

Layer	Kernel size	Stride	Padding	Output shape	Operations
Input	–	–	–	$[6, 30, 542, 221]$	–
Conv3D	3	1	1	$[9, 30, 542, 221]$	BN, ReLU, Dropout (d_r)
Conv3D	3	1	1	$[16, 30, 542, 221]$	BN, ReLU, Dropout (d_r)
Conv3D	3	1	1	$[32, 30, 542, 221]$	BN, ReLU, Dropout (d_r)
Conv3D	3	1	1	$[64, 30, 542, 221]$	BN, ReLU, Dropout (d_r)
Conv3D	3	1	1	$[128, 30, 542, 221]$	BN, ReLU, Dropout (d_r)
Conv3D	3	1	1	$[1, 30, 542, 221]$	–
Output	–	–	–	$[1, 30, 542, 221]$	–

3.3. *MuSt3Net two-phase training procedure*

The training procedure is structured into two distinct phases, summarized in Fig. 2. Unlike traditional data fusion approaches, which merge heterogeneous data sources instantaneously (Howard et al., 2024), we adopt a sequential strategy that progressively integrates different data modalities.

In the first phase, the network is trained to reconstruct 3D chlorophyll fields using MedBFM output as reference. At each weekly timestep, we extract k synthetic vertical profiles from the MedBFM chlorophyll field to approximate the typical sampling density of the BGC-Argo float. The model receives as input a 3D tensor of physical variables together with a sparse chlorophyll tensor constructed from these synthetic profiles, and it learns to reproduce the full MedBFM chlorophyll field. The loss is computed between the predicted and reference MedBFM fields, allowing the network to capture the spatial structure and variability of a dynamically consistent biogeochemical system.

Once the network has learned to emulate MedBFM fields, the second phase fine-tunes the model using chlorophyll profiles from BGC-Argo floats. At each timestep, float profiles are split into training and evaluation subsets. The input to the network consists of the full 3D numerical-model fields (identical to those used in the first phase) together with the sparse chlorophyll tensor constructed with the training-subset profiles. The model is then optimized by minimizing the loss between its predicted chlorophyll field and the independent evaluation profiles. The output consists of a 3D reconstructed chlorophyll field.

3.4. *Prediction confidence interval*

Beyond making predictions, it is essential to quantify uncertainty, especially when dealing with data from large and heterogeneous regions where measurement variability is high. For this task, we consider the confidence interval (Hosmer and Lemeshow, 1992), a widely used statistical measure that quantifies prediction uncertainty in different NN applications (Cortés-Ciriano and Bender, 2018; Salem et al., 2020; Schupbach et al., 2020). A practical approach to estimating predictive uncertainty is through deep ensembles (Pearce et al., 2018). These methods improve performance while naturally quantifying uncertainty by training multiple models independently. Since each model inherently deals with a slightly different optimization landscape (because of the stochasticity of the weights initialization), variations in their predictions serve as an uncertainty measure (Lakshminarayanan et al., 2017).

In our case, we generate ten predictions per input using an ensemble of independently trained neural networks. The standard deviation of these predictions quantifies confidence: a low standard deviation indicates higher confidence, while a high standard deviation suggests uncertainty due to data limitations, high input variability, or model complexity.

3.5. *Loss function*

The selection and design of an appropriate loss function play a crucial role in DL models, as they define the optimization objective and influence the model's ability to generalize to unseen data (Zhao et al., 2016). For MuSt3Net, the Root Mean Square Error (RMSE) (Hodson, 2022) is used as the loss function, measuring the difference between predicted and expected chlorophyll concentrations. In both phases, the RMSE is computed grid-wise, averaged across all depth levels. In the first phase, the RMSE is evaluated over the entire 3D domain, whereas in the second phase it is computed only at the grid locations where float observations are available. The same approach is used to compute the standard deviation, but substituting the mean with the standard-deviation operator. Moreover, two additional elements have been introduced to RMSE computation to better align with the chlorophyll reconstruction task.

First, a land-sea mask is applied to restrict loss computation to valid oceanic regions. This ensures that the optimization process remains unaffected by land pixels, where chlorophyll values are inherently zero, preventing artificial error minimization in these areas. The land-sea mask, already incorporated during input preprocessing, is also applied before computing the loss between the model output and the reference chlorophyll data.

Secondly, in the second training phase, the loss is computed between BGC-Argo observations and the corresponding temporal and spatial matching values of the 3D model reconstructed fields. Since BGC-Argo float data provide chlorophyll concentrations at specific spatial positions, a dedicated mask is generated to select these locations, ensuring that the optimization focuses exclusively on real in-situ data.

4. Experimental study

This section presents the experimental settings for the MuSt3Net architecture. The complete code for the reproducibility of the results presented in this paper is available at https://github.com/TeresaTonelli/MuSt3Net/tree/v1.4.0.

4.1. *Training*

We divided the dataset into three subsets: training, validation, and testing. The training set was used to optimize model parameters, while the validation set guided hyperparameter tuning and model selection (Pietropolli et al., 2024). The testing set evaluated the model's

Table 2
CNN hyperparameters used in our experiments.

Hyperparameter	Value
Epochs (Phase 1)	200
Epochs (Phase 2)	20
Learning rate	0.001 (both phases)
Hidden layers	6
Dropout rate	0.15
Activation function	ReLU
Kernel size	$3 \times 3 \times 3$
Trainable parameters	300 097

Table 3
Spatiotemporal discretization parameters.

Hyperparameter	Value
Latitude range	30°N –46°N
Longitude range	2°W–36°E
Depth range	0–300 m
Time range	2019–2021
Latitude resolution	8 km
Longitude resolution	8 km
Depth resolution	10 m
Time resolution	Weekly

performance on unseen data, assessing its generalization ability. The split was set to 80% for training, 10% for validation, and 10% for testing (Twomey et al., 1997).

The first phase ran for 200 epochs, while the second phase was limited to 20 epochs to mitigate overfitting, given the small dataset (< 250 samples) (Santos and Papa, 2022). We used the Rectified Linear Unit (ReLU) activation function in all hidden layers (Bai, 2022),we adopted an isotropic $3 \times 3 \times 3$ kernel to capture local spatial correlations without introducing an a priori directional bias and, for optimization, we used the Adam optimizer with a learning rate of 0.001 (Bock and Weiß, 2019). To further prevent overfitting, dropout regularization was applied with a probability of 0.15, following Roussillon et al. (2023). No layers were frozen in Phase 2 to avoid preserving biases inherited from the Phase 1 emulator. The hyperparameters used for training the MuSt3Net architectures are summarized in Table 2.

The training was performed on an NVIDIA A100-SXM GPU model with 64 GB of memory, leveraging CUDA 12.4 and PyTorch 2.4.0.

4.2. Experimental settings

The study domain covers the entire Mediterranean Sea, discretized at a horizontal resolution of $1/12°$ (about 8 km) and vertical resolution of 10 m from the surface down to 300 m depth. The dataset spans 2019–2021 at a weekly resolution, derived by averaging model outputs within each week. The hyperparameters for spatiotemporal discretization are summarized in Table 3.

The dataset is derived from MedBFM simulations and BGC-Argo float observations, both processed into 4D tensors (latitude, longitude, depth, and physical/biogeochemical variables). When multiple observations fall into the same spatial grid cell, their mean is assigned. The physical input variables include salinity, temperature, water velocity, and sea level anomaly, while chlorophyll is the only biogeochemical variable considered.

To align chlorophyll sampling between the two training phases, we downsample MedBFM data to match the lower density of float observations. Specifically, 200 MedBFM chlorophyll profiles are selected per week in the first phase, comparable to the ~ 50 float observations per week used in the second phase.

Seasonal variations in chlorophyll are accounted for when setting the mean-field, with values of 0.15 in summer and 0.075 in winter, reflecting increased biological activity in warmer months. The mean-field is used to fill missing data instead of assigning zero, as setting missing values close to the actual mean of the profile prevents introducing artificial biases, preserves the overall statistical distribution of the dataset and distinguishes them from land regions, commonly identified with zero (Zhang and Thorburn, 2022). The use of seasonal means to fill missing values avoids introducing artificial gradients and allows our model to learn the underlying dynamics independently, while still ensuring a consistent input field (Volpe et al., 2018).

4.3. A posterior validation analysis

To validate the MuSt3Net architecture, we conducted a thorough analysis of its performance in different geographic areas (NWM, TYR, SWM, ION, and LEV in Fig. 1) and across the four seasons: winter (JFM), spring (AMJ), summer (JAS), and fall (OND). While the MuSt3Net model is trained and tested on the entire dataset, this subdivision is only used to analyze the performance retrospectively to check whether the non-uniform geographical distribution and the natural variability in the chlorophyll field (e.g., depth and intensity of the deep chlorophyll maximum (DCM) in different Mediterranean regions) have an influence (Lavigne et al., 2015). In particular, the RMSE is calculated for the reconstructed profiles in each area and season to verify the presence of any bias in the accuracy of MuSt3Net in capturing the spatial and temporal variability of chlorophyll patterns in the Mediterranean Sea.

5. Results

This section reports the results of the MuSt3Net. We evaluate model performance using both quantitative metrics (e.g., RMSE) and qualitative ones (e.g., visualizations of the maps and profiles from the test set). To assess the effectiveness of our approach, we structure the analysis of MuSt3Net results around three central issues under investigation (RQ): **RQ1** Can our model, after the first phase, reconstruct realistic chlorophyll fields and successfully emulate MedBFM outputs? **RQ2** Does our model correctly integrate the BGC-Argo information after the second phase? **RQ3** Is our model able to capture the seasonal and regional variability of the chlorophyll field as reported in literature?

5.1. MuSt3Net 1^{st} phase: a MedBFM emulator

Fig. 3 shows the chlorophyll distribution computed by our model after the first training phase at different depth layer intervals, for a representative winter week (left panel) and for a summer week (right panel). To evaluate how well the MuSt3Net model emulates MedBFM, we compare its output with MedBFM predictions for the same weeks. The visual similarity between the two sets of maps confirms the capability of our model to reproduce the 3D chlorophyll field using only a limited number of input profiles. MuSt3Net captures both the west-to-east chlorophyll gradient both at the surface and in the subsurface layer (Basterretxea et al., 2018) and the expected seasonal differences (Lazzari et al., 2012). Notably, regional patterns, such as chlorophyll abundance in the Adriatic and Aegean Seas (Auricht et al., 2022), are well reconstructed, particularly in the surface layer.

Due to the higher chlorophyll variability of winter weeks, some differences between MuSt3Net and MedBFM chlorophyll field emerge: discrepancies are most pronounced in surface and coastal areas, which may be caused by the lower performance of CNNs typically observed at the border regions (Gupta et al., 2018). In general, the CNN-generated maps appear to have smoother patterns than those from MedBFM, reflecting the reduction of the horizontal resolution of CNN model. Nonetheless, the maps' similarities and the western–eastern chlorophyll gradient's prediction confirm the accuracy of the reconstruction computed by the first training phase.

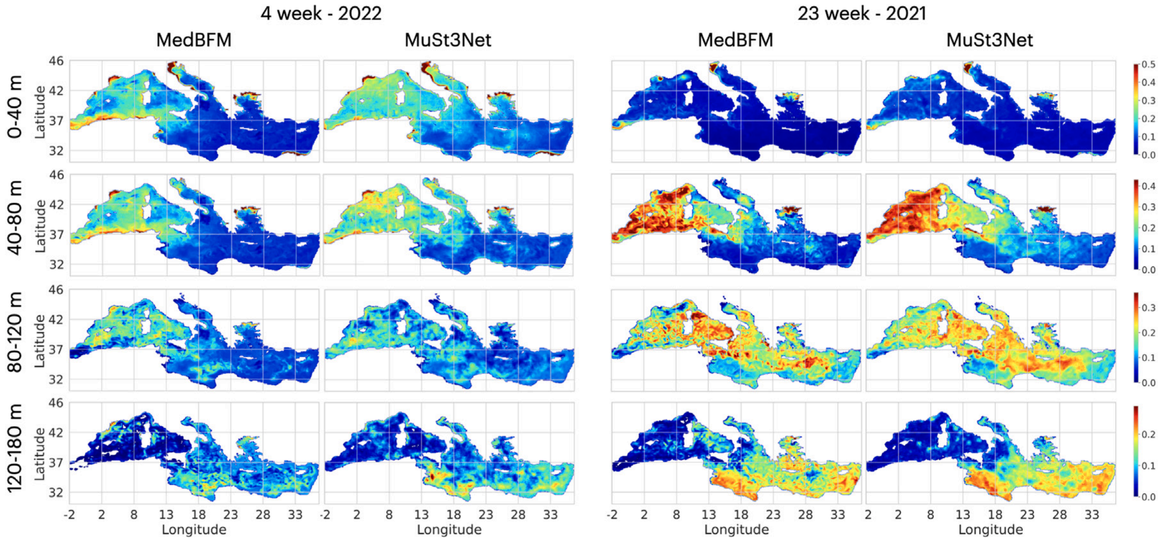


Fig. 3. Comparison of chlorophyll maps from MedBFM and first-phase MuSt3Net for a summer (week 23) and winter (week 4) week, averaged across depth layers: [0 m–40 m], [40 m–80 m], [80 m–120 m], and [120 m–180 m].

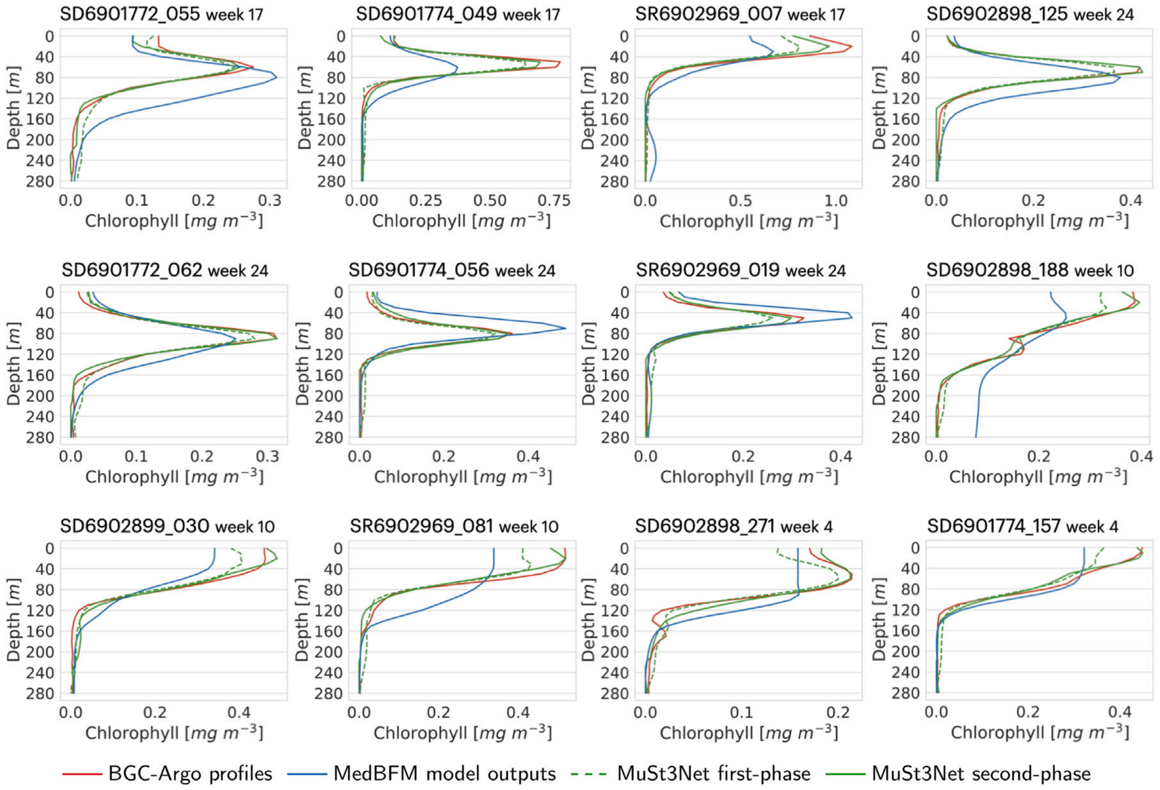


Fig. 4. Chlorophyll profiles for selected floats, including the corresponding WMO number and cycle to indicate location and time. The graphical legend represents different profile types. Profiles are taken from the test subset. The profiles refer to 2019 (17^{th}, 24^{th} week), 2020 (10^{th} week), and 2021 (4^{th} week)

5.2. MuSt3Net 2^{nd} phase: overview of BGC-argo floats integration

Fig. 4 shows a selection of 1D chlorophyll profiles predicted by MuSt3Net after the two-phase training. All profiles belong to the test set, meaning they were not used during training. To ensure spatial and temporal diversity in showing results, the profiles were selected from five distinct regions and across different seasons. Each plot displays the BGC-Argo float measurement (red), the MedBFM output (blue), the first-phase MuSt3Net prediction when applied to second-phase inputs, i.e. when MedBFM chlorophyll is replaced by BGC-Argo observations (dashed green), and the final MuSt3Net prediction after the second phase (solid green). This comparison illustrates how the model refines its predictions after integrating the information from BGC-Argo data.

For both the winter and summer weeks, MuSt3Net profiles provide a strong match with BGC-Argo float data, reproducing realistic values and profile shapes (e.g., correctly identifying the deep chlorophyll maximum (DCM) in summer). The model successfully captures seasonal variability, learning to distinguish between summer (e.g. SR6901772_062) and winter (e.g. SR6902969_081) profiles. This is evident in the improved alignment between the solid green line

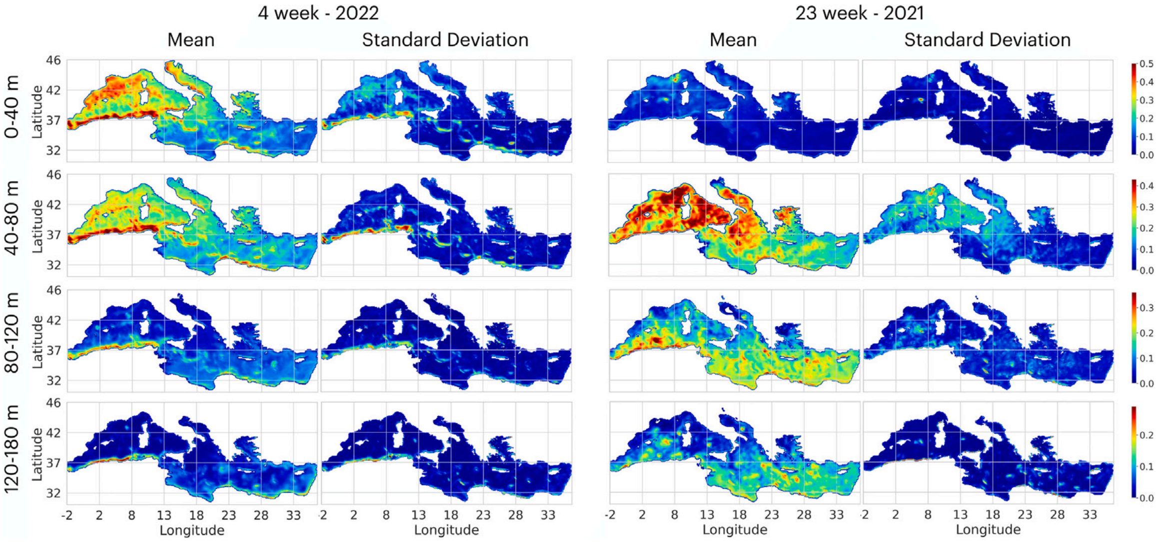


Fig. 5. Second-phase MuSt3Net chlorophyll mean and standard deviation maps for a summer (week 23) and winter (week 4) week, averaged across depth layers: [0 m–40 m], [40 m–80 m], [80 m–120 m], and [120 m–180 m].

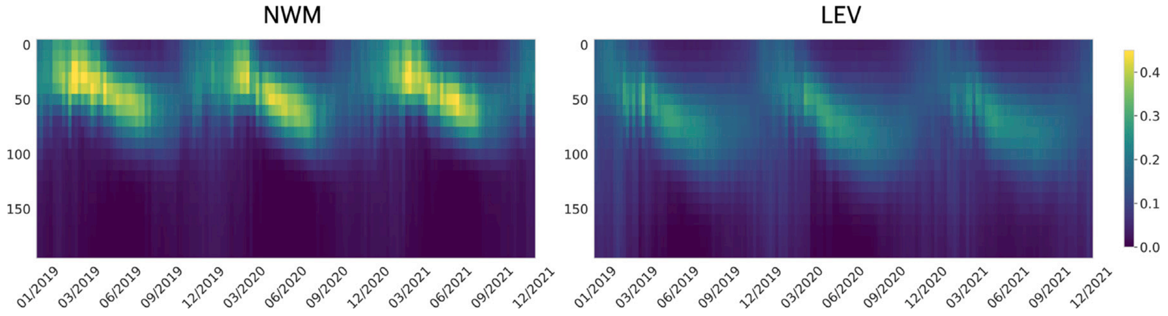


Fig. 6. Hovmöller diagram for the chlorophyll in two selected locations of the Mediterranean Sea: NWM (on the left) and LEV (on the right).

(final output) and the red line (BGC-Argo data), compared to the dashed green line (first-phase output), which confirms the ability of our model to integrate BGC-Argo data and refine its predictions. MedBFM profiles are generally slightly smoother than BGC-Argo observations, especially in winter, reflecting their nature as model-based approximations of reality. Since no data assimilation has been applied to MedBFM, its profiles diverge more from the BGC-Argo data than those predicted by MuSt3Net. Although trained to emulate MedBFM, the predictions of the first-phase MedBFM differ from the MedBFM profiles when applied to second-phase inputs. This is due to the change in input data: while the first training phase uses physical and chlorophyll fields from MedBFM, the second phase replaces chlorophyll MedBFM inputs with BGC-Argo float data, which are generally higher than the modeled values. As a consequence, the first-phase MedBFM produces outputs that deviate from the original MedBFM behavior, explaining the differences between the blue and dashed green lines. These dashed profiles are not meant to serve as a numerical baseline but rather to illustrate the contrast before and after the fusion step.

The profiles provide a qualitative assessment of the effectiveness of the proposed data fusion method. To complement this with a quantitative evaluation, we compute the RMSE between the reconstructed profiles and BGC-Argo float data across different seasons and geographical regions (Table 4), together with the corresponding standard deviations. The RMSE values between BGC-Argo float data and MuSt3Net reconstructions, separated into training and test sets, vary from $0.05\,\mathrm{mg\,m^{-3}}$ to $0.12\,\mathrm{mg\,m^{-3}}$, with lower errors in summer and fall, and higher ones in winter and spring. The increased RMSE in winter and spring is likely due to the presence of surface chlorophyll blooms, especially in western sub-basins, which increase the dynamic range of chlorophyll values, thus amplifying reconstruction error (Teruzzi et al., 2021; Cossarini et al., 2019). To better evaluate seasonal variability, we compute the RMSE separately for profiles affected by two key temporal features: surface blooms and deep DCM. Since these phenomena occur during specific periods of the year, we evaluated the model performance on the corresponding months: February–April for surface blooms and May–September for deep DCM. The RMSE is approximately $0.101\,\mathrm{mg\,m^{-3}}$ for surface bloom and $0.075\,\mathrm{mg\,m^{-3}}$ for DCM, confirming the ability of our model to reconstruct these features, consistent with the corresponding seasonal RMSEs in Table 4.

The differences between train and test RMSE are more accentuated in winter (test RMSE increase of 46%) than in summer (test RMSE increase of 10%), confirming the higher challenges in predicting winter chlorophyll dynamics than summer ones.

Table 5 reports the RMSE between BGC-Argo float data and MuSt3Net reconstructions across different geographical areas, separated into training and test sets. In all regions, the RMSE remains below $0.1\,\mathrm{mg\,m^{-3}}$, consistent with the expected accuracy of biogeochemical numerical models, such as the one proposed in Salon et al. (2019), which provide a valid chlorophyll reconstruction in the Mediterranean Sea. The highest RMSEs are observed in the western sub-basins (NWM,

Table 4
Error metrics (averaged over depths) between BGC-Argo float measurements and reconstructed values from the MuSt3Net second phase, computed on both train and test sets across different seasons.

Season	RMSE $[\mathrm{mg\,m^{-3}}]$		MAE $[\mathrm{mg\,m^{-3}}]$		BIAS $[\mathrm{mg\,m^{-3}}]$		r	
	Train	Test	Train	Test	Train	Test	Train	Test
Winter	0.065 ± 0.012	0.095 ± 0.009	0.060	0.066	−0.023	−0.025	0.99	0.99
Spring	0.099 ± 0.015	0.124 ± 0.017	0.056	0.064	−0.016	−0.021	0.97	0.96
Summer	0.071 ± 0.010	0.078 ± 0.012	0.043	0.045	−0.002	−0.003	0.99	0.98
Fall	0.055 ± 0.008	0.049 ± 0.005	0.034	0.031	−0.001	0.001	0.99	0.99

Table 5
Error metrics (averaged over depths) between BGC-Argo float measurements and reconstructed values from the MuSt3Net second phase, computed on both train and test sets across different geographic areas.

Season	RMSE $[\mathrm{mg\,m^{-3}}]$		MAE $[\mathrm{mg\,m^{-3}}]$		BIAS $[\mathrm{mg\,m^{-3}}]$		r	
	Train	Test	Train	Test	Train	Test	Train	Test
NWM	0.086 ± 0.029	0.096 ± 0.032	0.049	0.059	−0.015	−0.024	0.98	0.97
SWM	0.098 ± 0.031	0.086 ± 0.034	0.059	0.050	−0.027	−0.019	0.99	0.99
TYR	0.094 ± 0.038	0.099 ± 0.042	0.063	0.063	−0.028	−0.025	0.97	0.94
ION	0.053 ± 0.020	0.055 ± 0.019	0.043	0.044	−0.003	−0.003	0.98	0.99
LEV	0.065 ± 0.018	0.066 ± 0.021	0.045	0.046	−0.006	−0.004	0.99	0.99

SWM, TYR), where chlorophyll variability is greater (NWM) and BGC-Argo float coverage is more limited (SWM and TYR). As shown in Fig. 1, half of the available BGC-Argo profiles are concentrated in the Ionian and Levantine Seas, where reconstruction errors are correspondingly lower—down to $0.05\,\mathrm{mg\,m^{-3}}$ in the Ionian basin. Nonetheless the different approach's techniques, both western and eastern sub-basins yield RMSE values in line with previous studies (Salon et al., 2019; Pietropolli et al., 2024).

Importantly, training and test RMSE values are similar across all regions. Indeed, the increase of RMSE from training and test is from 4% (eastern basins) to 8% (western basins), indicating the absence of overfitting.

To strengthen our analysis, we complement the RMSE with the computation of additional performance metrics, reported in Tables 4 and 5. In particular, we compute the Mean Absolute Error (MAE), the bias (BIAS) and the Pearson Correlation Coefficient (r). The MAE evaluates the general reliability of the model, providing a robust and interpretable measure of the average magnitude of prediction errors while being less sensitive to outliers. The BIAS detects the systematic tendency of the model to overestimate or underestimate the observations, while r quantifies the strength of the linear relationship between predicted and observed values.

Consistent with the RMSE results, the MAE indicates larger errors in winter and spring seasons, as well as greater challenges in integrating profiles from underrepresented regions, such as the Tyrrhenian Sea. The bias analysis reveals that our proposed integration tends to slightly underestimate the observations from BGC-Argo floats. This behavior is expected, given that the BGC-Argo measurements are generally higher than the corresponding values from the MedBFM dataset. We compute the Pearson correlation coefficient to evaluate the temporal and spatial coherence between the CNN-integrated fields and the BGC-Argo float observations. However, as is common in reconstruction tasks, this metric provides limited insight. Since the reconstructed fields are designed to closely replicate the true data distribution, the relationship between the two is expected to be approximately linear, reducing the discriminative value of the correlation coefficient.

5.3. MuSt3Net 2^{nd} phase: a 3D model data fusion

Fig. 5 shows the final chlorophyll distribution predicted by MuSt3Net at different depth layers, for a winter week (left panel) and a summer week (right panel). For each week, we report the ensemble mean (left column) and the corresponding standard deviation (right column), computed over models trained with different weight initializations (Section 3.4). Assuming an average RMSE of approximately $0.07\,\mathrm{mg\,m^{-3}}$, we can consider MuSt3Net predictions reliable when ensemble standard deviations fall below this threshold. These maps reflect the output of the two-phase training procedure and the integration of BGC-Argo float data into a 3D framework.

The low standard deviation across most regions confirms the stability of the reconstruction, while higher variability is mostly confined to surface and coastal areas or regions with sparse BGC-Argo coverage.

The similarity between chlorophyll spatial patterns in both training phases (Figs. 3 and 5) confirms the effectiveness of the two-phase training procedure in integrating ARGO float data while preserving spatial consistency. The higher chlorophyll concentration in the upper layers ([0 m–40 m], [40 m–80 m]) during winter, compared to the first training phase (Fig. 3), reflect the typical surface bloom and are driven by the elevated surface values in BGC-Argo data. Being real-time observations, BGC-Argo profiles capture biogeochemical variability sometimes missed by the MedBFM approximation. Conversely, the lower chlorophyll values in deeper layers are due to a stronger vertical decrease in BGC-Argo profiles compared to MedBFM, as shown in Fig. 4.

The low standard deviation across ensemble members in most of the areas confirms the stability of the training and the limited sensitivity to weight initialization (Pearce et al., 2018). In ensemble setups, low variability indicates that the model consistently converges toward similar solutions, increasing confidence in prediction robustness. Higher standard deviations appear mostly in coastal areas, where the absence of BGC-Argo chlorophyll values on land and the presence of domain borders make prediction more challenging (Wang et al., 2024).

5.4. Temporal evolution of the chlorophyll field

Fig. 6 shows Hovmöller diagrams representing the temporal evolution of chlorophyll reconstructed by MuSt3Net. These plots assess the model's ability to capture temporal chlorophyll patterns, in addition to the spatial structures discussed in previous sections. To highlight contrasting dynamics and better underline the chlorophyll differences between sub-basins, without considering lands and coastal misleading contributions, we report the diagrams of the two areas (1.5° x 1.5° wide) in two subregions (see red crosses in Fig. 1) with the most different behavior: the NWM (left) and the LEV (right).

Fig. 6 shows that the temporal behavior of chlorophyll profiles is consistent with patterns reported in the literature, with surface blooms in winter and a reliable DCM in summer. Higher chlorophyll concentrations, indicated by more intense yellow colors, shift in depth over time: they appear near the surface during winter months and between [60 m – 100 m] in late spring and summer (Lazzari et al., 2012; Salon et al., 2019).

Both diagrams correctly reproduce typical chlorophyll dynamics, including the higher concentrations in the western basin (NWM) compared to the eastern one (LEV), as indicated by the more prominent yellow peaks in the left panel. The DCM cycle is accurately captured in both regions: in the NWM Sea, it stabilizes around 70 m, while in the LEV it forms deeper, around 90 m—both values consistent with those reported in the literature (Lavigne et al., 2015).

6. Discussions

In this work, we propose a new approach to data fusion through DL, exploiting NN capabilities, such as faster computation and high prediction accuracy, to integrate multiple data sources with an effectiveness that spreads throughout the entire 3D Mediterranean basin.

6.1. Deep learning for model data fusion in oceanography

The use of DL for data fusion is an interesting and still open research area, especially in oceanography, where data sources are diverse in availability, structure, and density (Brajard et al., 2021; Cossarini et al., 2019). Our work was motivated by the consideration that integrating multiple data sources – often characterized by differences in accuracy and structural characteristics – can improve reconstruction quality through the combination of complementary information that single data types cannot capture (e.g.different ocean observing systems such as Argo profiles) (Anderson et al., 2009). Managing such heterogeneity is a growing area of investigation in DL, where leveraging diverse data types enables models to describe complex phenomena better, improve prediction quality and realism across domains (Kline et al., 2022; Stahlschmidt et al., 2022). The comparison shown in Fig. 4 exemplifies how this approach, named Multimodal Deep Learning (MDL) (Gao et al., 2020; Ngiam et al., 2011), provides superior performance in reconstruction and prediction tasks compared to standard DL approaches (Kline et al., 2022; Stahlschmidt et al., 2022). Indeed, our multiple-phase training strategy is commonly used to exploit data-specific features in imaging (Guo et al., 2019). By correcting the MuSt3Net (first-phase), our sequential training strategy corrects profile discrepancies from earlier phases (Friedman, 2001) and retains previously acquired knowledge (Walach and Wolf, 2016). We implement this approach using a 3D-CNN that merges deterministic model outputs and observational data, in line with recent DL-based data fusion strategies (Barthélémy et al., 2022; Brajard et al., 2021; Gao et al., 2020; Howard et al., 2024). Rather than substituting observational data directly into model outputs, we introduce an integration step where data are split based on structural characteristics, preserving intrinsic information and preventing misinterpretation. For instance, directly substituting BGC-Argo float observations into the 3D MedBFM field – without separating them from model outputs into distinct training phases – could lead the CNN to misinterpret the vertical profiles of float data as discontinuities, resulting in incorrect integration. This approach, known as gradient boosting (Friedman, 2001; Natekin and Knoll, 2013), outperforms single-phase training methods (Emami and Martínez-Muñoz, 2023) and favors the use of simpler architectures, such as CNNs (Walach and Wolf, 2016). Furthermore, our sequential training strategy can be seen as a form of transfer learning (Weiss et al., 2016), a paradigm that shows to outperform traditional methods in scenarios where data heterogeneity and inhomogeneity require finer modeling (Hattula et al., 2023; Salehi et al., 2023). Oceanography is a promising yet underexplored field for transfer learning, due to the diversity of data sources and the inhomogeneity of data. To date, few works have applied transfer learning in this domain, and, to our knowledge, none have addressed chlorophyll reconstruction in a 3D domain using this approach. In Zhang et al. (2024), transfer learning is used to improve the sea surface height fields through the fusion of heterogeneous sea surface temperature data, and it is implemented through a GAN architecture. In Zhang et al. (2022), a transfer learning model, described as a fully connected neural network, is employed to retrieve the amplitude of internal waves from satellite images. A similar structure is adopted in Pietropolli et al. (2022), but with a different architecture (GANs instead of CNNs) and domain (2D instead of 3D). Modeling the 3D chlorophyll field is challenging due to multiple physical and biogeochemical factors acting both at the surface (Lazzari et al., 2012) and through the water column (Lavigne et al., 2015). Most DL approaches focus on 2D spatial domains, as satellite observations – the most available observation data source – are typically limited to surface data. Only a few studies have addressed the reconstruction of full 3D oceanic fields (Na et al., 2022; Sammartino et al., 2025), exploring architectures such as Multi-Layer Perceptrons (MLPs) (Sammartino et al., 2020) and CNNs (Bonino et al., 2024). CNNs are particularly suited for this task, as they have demonstrated strong performance in 3D field reconstruction (Cheng et al., 2016), feature extraction (Zhou et al., 2016), image super-resolution (Yatheendradas and Kumar, 2022), ocean forecasting (Hao et al., 2023; Zhou et al., 2021), transfer learning (Panwar et al., 2020), and model interpretability (Roussillon et al., 2023; Wang and Li, 2023; Wang et al., 2023).

In this work, CNNs are selected for their ability to generate spatially coherent 3D predictions, overcoming the limitations of MLPs (Pietropolli et al., 2023). While MLPs provide point-wise predictions often suffering from artificial discontinuities (Pietropolli et al., 2024; Skákala et al., 2023), CNNs exploit local spatial dependencies, producing smoother and more realistic outputs (Pietropolli et al., 2025; Pintelas and Pintelas, 2022; Pistellato et al., 2021). This property is particularly valuable in oceanographic applications.

6.2. Comparison with other data assimilation methods

Traditionally, the integration of BGC-Argo data – and vertical profiles more broadly – into biogeochemical models is achieved through DA techniques (Cossarini et al., 2019; Skakala et al., 2021; Verdy and Mazloff, 2017). These methods enable the correction of modeled vertical dynamics even at locations distant from the profiles. Our method represents an example of a model data fusion, where the background state (in our case, first-phase training) is updated by the second training based on BGC-Argo observations. Given the 3D structure of CNN architectures, the information is nonlinearly propagated across the domain. Our approach uses a CNN to integrate data, an operation that is traditionally executed by statistical operators. The use of DL architecture to perform data fusion is supported by both mathematical theory and experimental results (Bocquet et al., 2020). From a theoretical point of view, there are several similarities between the DL and DA variational approach, such as the equivalence between the variational DA's cost function and the CNN's loss function (Abarbanel et al., 2018), the employment of gradient descent techniques to minimize the cost function (or loss function respectively) (Abarbanel et al., 2018; Bocquet et al., 2020) and the mathematical identity between the adjoint method that calculates gradients in DA and the CNN's backpropagation (Hsieh and Tang, 1998). Compared to traditional DA (Law et al., 2015) and data fusion approaches (Brajard et al., 2021; Howard et al., 2024), our approach performs both an emulation and a data fusion, exploiting the same CNN architecture. Traditional data assimilation integrates observations to change the initial condition of a numerical model to make it more compatible with reality; it so generates a cycle, in which model outputs are combined with observations to provide initial conditions that the numerical model will use to predict the subsequent time step. Both the numerical model and the DA are generally very complex procedures from a computational point of view. Surrogate models are traditionally used to simplify the model step, while DA techniques are used to enhance the observations' integration. Our approach proposes to develop the same CNN as both emulator and data integrator: in this way, the information gained in each step is preserved and transferred to the other, also enhancing and supporting the development of the two-phase training procedure.

6.3. Computational time

A direct performance comparison with DA methods is not straightforward–since our approach first emulates the deterministic model to generate a background state for the second phase–and it is beyond the scope of the present work. Nevertheless, the proposed approach offers a potential key advantage in computational efficiency (Šíma and Orponen, 2003). This is made possible by modern, user-friendly open-source software and advances in computing cluster architectures (Yan et al., 2024). Indeed, as multiple DL approaches (Bengio et al., 2015), our two-phase training procedure computes the chlorophyll 3D field in a lower computational time (order of seconds) with respect to traditional DA approaches (order of hours) once trained. Compared to standard DA approaches, the saved computational time can be spent on MuSt3Net internal improvements.

6.4. Limitations and future works

Our two-phase training strategy comes with some limitations. The sparse spatial coverage of BGC-Argo floats constrains the effective embedding of observations, making it challenging to reconstruct chlorophyll dynamics in under-sampled regions. The first training phase helps mitigate this issue by allowing the network to learn large-scale chlorophyll patterns from the numerical model; however, areas with very limited observations, (e.g. SWM, TYR and coastal regions), can remain reconstructed with lower accuracy. A denser float dataset would likely reduce these regional discrepancies. The uneven distribution of BGC-Argo profiles also leads to under-represented sub-basins (such as the Tyrrhenian Sea), where our reconstruction shows higher RMSE values (Table 5). Despite these limitations, the method preserves mesoscale variability and reproduces realistic vertical structures, as shown in Figs. 4 and 5. Another limitation is the absence of explicit temporal modeling: time enters the network only through independent samples at the weekly frequency. Unlike traditional DA approaches, our framework does not encode temporal dependencies directly. Nevertheless, the temporal evolution of chlorophyll remains coherent after float integration (Fig. 6), suggesting that spatial patterns learned from the model still guide the reconstruction. Since this work represents a first attempt at a two-phase fusion strategy, we focused on the integration procedure rather than on temporal architectures. Exploring these extensions is a natural next step.

This work represents a baseline for future multiple training phases model data fusion approaches, and different adjustments can be implemented to make it suitable for more specific purposes. The integration of satellite data, which can be operated simply through the addition of a third training phase, could improve the prediction of surface phenomena, such as the surface bloom (Vidussi et al., 2000). In environmental sciences, the estimation of the prediction uncertainties is important to provide information about the reconstructed phenomenon; thus, the introduction of some uncertainty quantification (UQ) techniques could improve the reconstruction analysis (Tripathy and Bilionis, 2018). The UQ techniques model the uncertainties of the prediction, estimating the optimal parameters of reconstructed data distribution (Haynes et al., 2023). Embedding our loss function with a UQ technique, such as parametric distributional prediction (PDP) (Barnes and Barnes, 2021) or ensemble prediction (EP) (Lee and Baldick, 2013), we compute both the chlorophyll 3D field and an estimation of its uncertainty, thus providing another qualitative measure for 3D field reconstruction. Our MuSt3Net architecture does not include the time, limiting the ability to recognize temporal behavior for all the geographical areas. A combination of CNN and networks that encounter time, such as recurrent neural networks (RNN) (Sherstinsky, 2020), could improve the reconstruction quality on a spatiotemporal scale (Lazzari et al., 2021); an interesting example, widely applied in environmental sciences, is represented by convolutional long short-term memory (ConvLSTMs) (Wang and Li, 2023).

7. Conclusions

We introduce a novel two-phase CNN approach for reconstructing biogeochemical ocean fields. The first phase emulates the output of a biogeochemical model, while the second integrates real observational data. This decoupled training strategy preserves the intrinsic characteristics of each data source, preventing misinterpretation during integration and leveraging the capacity of the CNN to capture spatial correlation patterns without inducing cross-source biases. Applied to chlorophyll reconstruction in the Mediterranean Sea, our method integrates process-based model output and BGC-Argo profiles and produces reliable 3D fields that capture characteristic spatial structures and temporal dynamics, with error levels in line with traditional DA. The compact architecture ensures fast training and inference while maintaining high reconstruction quality and can be exported to other variables from BGC-Argo floats. Finally, its sequential structure enables easy extension, as additional data types can be integrated by adding a dedicated training phase.

CRediT authorship contribution statement

Teresa Tonelli: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Gianpiero Cossarini:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Gloria Pietropolli:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Luca Manzoni:** Conceptualization of this study, Methodology, Project administration, Supervision, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the NECCTON project, which has received funding from Horizon Europe RIA under grant agreement No 101081273. This study has been conducted using E.U. Copernicus Marine Service Information.

Appendix. Extended results

A.1. Losses evolution across epochs

Fig. 7 illustrates the evolution of the training and test losses across epochs for both the first (top row) and second (bottom row) training phases. In both phases, the loss curves for the training and test sets decrease consistently, indicating the absence of overfitting. The smoothness of the loss trajectories further suggests stable learning dynamics and the ability of the CNN to progressively improve its performance over epochs. Overall, the loss curves demonstrate that the model successfully converges while maintaining good generalization.

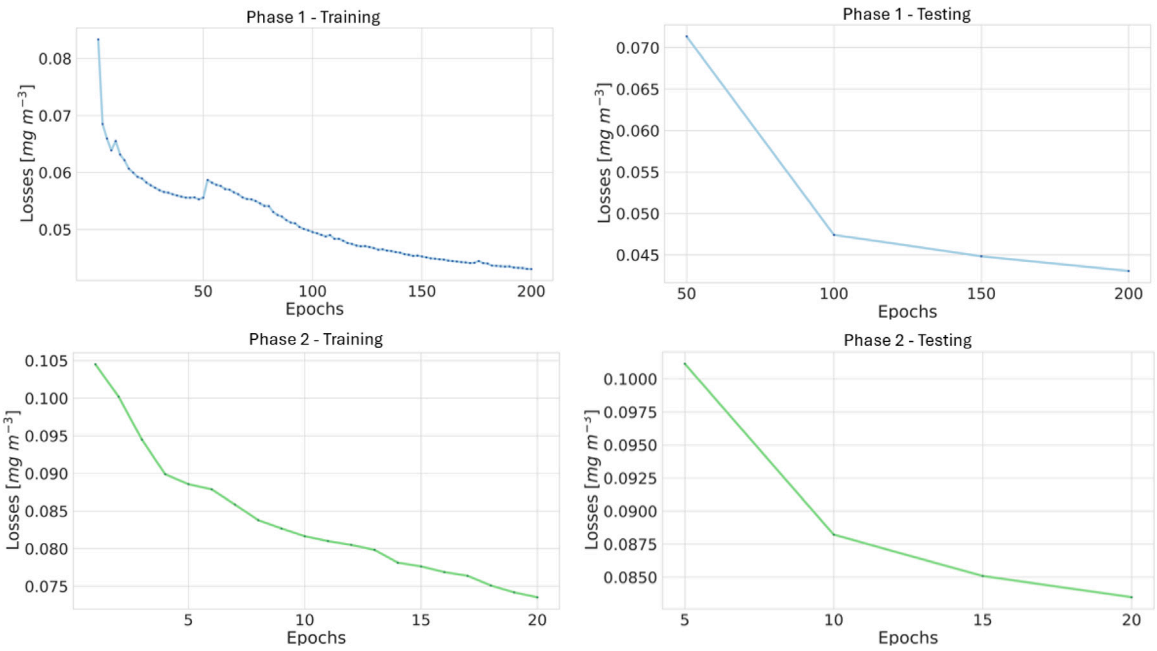


Fig. 7. Loss curves for Phase 1 (top) and Phase 2 (bottom). Training and testing losses are shown in the left and right panels, respectively.

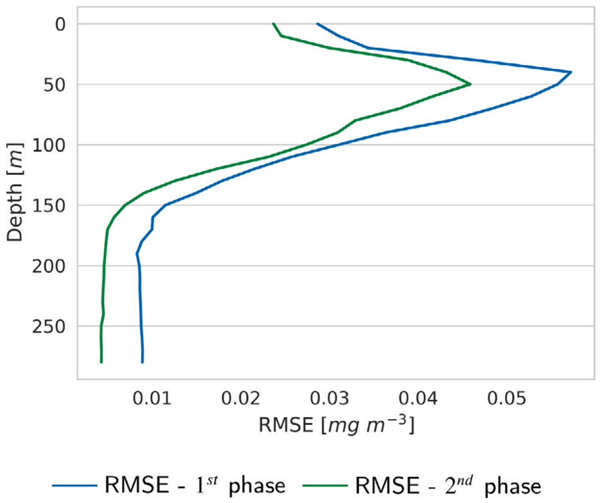


Fig. 8. RMSE variation across depth levels, computed on the test set for the first (blue line) and the second phase (green line).

A.2. RMSE across depth layers: A comparative analysis of 1^{st} and 2^{nd} phase

Fig. 8 shows the variation of RMSE across depth layers, comparing the RMSE of the 1^{st} and 2^{nd} phase on the test set of observations. As expected, the RMSE in the second phase – which is designed to integrate BGC-Argo float data – is lower across all depth layers compared to the RMSE between the BGC-Argo floats and the corresponding profiles predicted by the emulator, which is trained solely to emulate the MedBFM data rather than integrate the floats. Moreover, the evolution of RMSE across depths presents higher values in superficial layers and near DCM, compatibly with the complexity in modeling these key features.

Data availability

Data will be made available on request.

References

Abarbanel, H.D., Rozdeba, P.J., Shirman, S., 2018. Machine learning: Deepest learning as statistical data assimilation problems. Neural Comput. 30 (8), 2025–2055.

Amadio, C., Teruzzi, A., Feudale, L., Bolzon, G., Di Biagio, V., Lazzari, P., Álvarez, E., Coidessa, G., Salon, S., Cossarini, G., 2023. Mediterranean Quality checked BGC-Argo 2013–2022 dataset. http://dx.doi.org/10.5281/zenodo.10391759.

Amadio, C., Teruzzi, A., Pietropolli, G., Manzoni, L., Coidessa, G., Cossarini, G., 2024. Combining neural networks and data assimilation to enhance the spatial impact of Argo floats in the Copernicus Mediterranean biogeochemical model. Ocean. Sci. 20 (3), 689–710.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Avellano, A., 2009. The data assimilation research testbed: A community facility. Bull. Am. Meteorol. Soc. 90 (9), 1283–1296.

Argo, 2000. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). http://dx.doi.org/10.17882/42182, Last visited on August 2022.

Asher, M.J., Croke, B.F., Jakeman, A.J., Peeters, L.J., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51 (8), 5957–5973.

Auricht, H., Mosley, L., Lewis, M., Clarke, K., 2022. Mapping the long-term influence of river discharge on coastal ocean chlorophyll-a. Remote. Sens. Ecol. Conserv. 8 (5), 629–643.

Bai, Y., 2022. RELU-function and derived function review. In: SHS Web of Conferences. Vol. 144, EDP Sciences, p. 02006.

Baldi, P., Sadowski, P.J., 2013. Understanding dropout. Adv. Neural Inf. Process. Syst. 26.

Barnes, E.A., Barnes, R.J., 2021. Controlled abstention neural networks for identifying skillful predictions for regression problems. J. Adv. Model. Earth Syst. 13 (12), e2021MS002575.

Barthélémy, S., Brajard, J., Bertino, L., Counillon, F., 2022. Super-resolution data assimilation. Ocean. Dyn. 72 (8), 661–678.

Basterretxea, G., Font-Muñoz, J.S., Salgado-Hernanz, P.M., Arrieta, J., Hernández-Carrasco, I., 2018. Patterns of chlorophyll interannual variability in Mediterranean biogeographical regions. Remote Sens. Environ. 215, 7–17.

Bengio, E., Bacon, P.-L., Pineau, J., Precup, D., 2015. Conditional computation in neural networks for faster models. arXiv preprint arXiv:1511.06297.

Bittig, H.C., Maurer, T.L., Plant, J.N., Schmechtig, C., Wong, A.P., Claustre, H., Trull, T.W., Udaya Bhaskar, T., Boss, E., Dall'Olmo, G., et al., 2019. A BGC-Argo guide: Planning, deployment, data handling and usage. Front. Mar. Sci. 6, 502.

Bittig, H., Wong, A., Plant, J., 2022. BGC-Argo synthetic profile file processing and format on Coriolis GDAC.

Bjorck, N., Gomes, C.P., Selman, B., Weinberger, K.Q., 2018. Understanding batch normalization. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 31, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper_files/paper/2018/file/36072923bfc3cf47745d704feb489480-Paper.pdf.

Bock, S., Weiß, M., 2019. A proof of local convergence for the Adam optimizer. In: 2019 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Bocquet, M., 2023. Surrogate modeling for the climate sciences dynamics with machine learning and data assimilation. Front. Appl. Math. Stat. 9, 1133226.

Bocquet, M., Brajard, J., Carrassi, A., Bertino, L., 2020. Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. arXiv preprint arXiv:2001.06270.

Bonino, G., Galimberti, G., Masina, S., McAdam, R., Clementi, E., 2024. Machine learning methods to predict sea surface temperature and marine heatwave occurrence: a case study of the Mediterranean Sea. Ocean. Sci. 20 (2), 417–432.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2021. Combining data assimilation and machine learning to infer unresolved scale parametrization. Phil. Trans. R. Soc. A 379 (2194), 20200086.

Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. Wiley Interdiscip. Rev.: Clim. Chang. 9 (5), e535.

Castanedo, F., 2013. A review of data fusion techniques. Sci. World J. 2013 (1), 704504.

Cetinic, E., Lipic, T., Grgic, S., 2018. Fine-tuning convolutional neural networks for fine art classification. Expert Syst. Appl. 114, 107–118.

Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (12), 7405–7415.

Coppini, G., Clementi, E., Cossarini, G., Salon, S., Korres, G., Ravdas, M., Lecci, R., Pistoia, J., Goglio, A.C., Drudi, M., et al., 2023. The Mediterranean Forecasting System–Part 1: Evolution and performance. Ocean. Sci. 19 (5), 1483–1516.

Cortés-Ciriano, I., Bender, A., 2018. Deep confidence: a computationally efficient framework for calculating reliable prediction errors for deep neural networks. J. Chem. Inf. Model. 59 (3), 1269–1281.

Cossarini, G., Feudale, L., Teruzzi, A., Bolzon, G., Coidessa, G., Solidoro, C., Di Biagio, V., Amadio, C., Lazzari, P., Brosich, A., et al., 2021. High-resolution reanalysis of the Mediterranean Sea biogeochemistry (1999–2019). Front. Mar. Sci. 8, 741486.

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., d'Ortenzio, F., 2019. Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea. Ocean. Model. 133, 112–128.

d'Ortenzio, F., Lavigne, H., Besson, F., Claustre, H., Coppola, L., Garcia, N., Laës-Huon, A., Le Reste, S., Malardé, D., Migon, C., et al., 2014. Observing mixed layer depth, nitrate and chlorophyll concentrations in the northwestern Mediterranean: A combined satellite and NO3 profiling floats experiment. Geophys. Res. Lett. 41 (18), 6443–6451.

Ducournau, A., Fablet, R., 2016. Deep learning for ocean remote sensing: An application of convolutional neural networks for super-resolution on satellite-derived SST data. In: 2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing. PRRS, IEEE, pp. 1–6.

Emami, S., Martínez-Muñoz, G., 2023. A gradient boosting approach for training convolutional and deep neural networks. IEEE Open J. Signal Process. 4, 313–321.

Escudier, R., Clementi, E., Cipollone, A., Pistoia, J., Drudi, M., Grandi, A., Lyubartsev, V., Lecci, R., Aydogdu, A., Delrosso, D., et al., 2021. A high resolution reanalysis for the Mediterranean Sea. Front. Earth Sci. 9, 702285.

Fennel, K., Mattern, J.P., Doney, S.C., Bopp, L., Moore, A.M., Wang, B., Yu, L., 2022. Ocean biogeochemical modelling. Nat. Rev. Methods Prim. 2 (1), 76.

Forrester, A.I., Sóbester, A., Keane, A.J., 2007. Multi-fidelity optimization via surrogate modelling. Proc. R. Soc. A: Math. Phys. Eng. Sci. 463 (2088), 3251–3269.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Gao, J., Li, P., Chen, Z., Zhang, J., 2020. A survey on deep learning for multimodal data fusion. Neural Comput. 32 (5), 829–864.

Gettelman, A., Geer, A.J., Forbes, R.M., Carmichael, G.R., Feingold, G., Posselt, D.J., Stephens, G.L., Van den Heever, S.C., Varble, A.C., Zuidema, P., 2022. The future of Earth system prediction: Advances in model-data fusion. Sci. Adv. 8 (14), eabn3488.

Gregg, W.W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. J. Mar. Syst. 69 (3–4), 205–225.

Guo, Z., Li, X., Huang, H., Guo, N., Li, Q., 2019. Deep learning-based image segmentation on multimodal medical imaging. IEEE Trans. Radiat. Plasma Med. Sci. 3 (2), 162–169.

Gupta, H., Jin, K.H., Nguyen, H.Q., McCann, M.T., Unser, M., 2018. CNN-based projected gradient descent for consistent CT image reconstruction. IEEE Trans. Med. Imaging 37 (6), 1440–1453.

Gutknecht, E., Bertino, L., Brasseur, P., Ciavatta, S., Cossarini, G., Fennel, K., Ford, D., Grégoire, M., Lavoie, D., et al., 2022. Biogeochemical modelling. In: Implementing Operational Ocean Monitoring and Forecasting Systems. GOOS-275, IOC-UNESCO publication, pp. 247–306.

Hall, D.L., Llinas, J., 1997. An introduction to multisensor data fusion. Proc. IEEE 85 (1), 6–23.

Hao, P., Li, S., Song, J., Gao, Y., 2023. Prediction of sea surface temperature in the South China Sea based on deep learning. Remote. Sens. 15 (6), 1656.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. vol. 2, Springer.

Hattula, E., Zhu, L., Raninen, J., Oksanen, J., Hyyppä, J., 2023. Advantages of using transfer learning technology with a quantative measurement. Remote. Sens. 15 (17), 4278.

Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K., Ebert-Uphoff, I., 2023. Creating and evaluating uncertainty estimates with neural networks for environmental-science applications. Artif. Intell. the Earth Syst. 2 (2), 220061.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geosci. Model. Dev. Discuss. 2022, 1–10.

Hosmer, D.W., Lemeshow, S., 1992. Confidence interval estimation of interaction. Epidemiology 452–456.

Hoteit, I., Luo, X., Bocquet, M., Kohl, A., Ait-El-Fquih, B., 2018. Data assimilation in oceanography: Current status and new directions. New Front. Oper. Ocean. 465–512.

Howard, L.J., Subramanian, A., Hoteit, I., 2024. A machine learning augmented data assimilation method for high-resolution observations. J. Adv. Model. Earth Syst. 16 (1), e2023MS003774.

Hsieh, W.W., Tang, B., 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. Bull. Am. Meteorol. Soc. 79 (9), 1855–1870.

Kantha, L.H., Clayson, C.A., 2000. Numerical Models of Oceans and Oceanic Processes. vol. 66, Elsevier.

Karagiorgos, J., Vervatis, V., Sofianos, S., 2025. Ocean chlorophyll feedback in a coupled ocean-atmosphere model for the Mediterranean and Black Seas. J. Geophys. Res.: Ocean. 130 (2), e2024JC021985.

Kartal, S., Sekertekin, A., 2022. Prediction of MODIS land surface temperature using new hybrid models based on spatial interpolation techniques and deep learning models. Environ. Sci. Pollut. Res. 29 (44), 67115–67134.

Kleijnen, J.P., 2009. Kriging metamodeling in simulation: A review. European J. Oper. Res. 192 (3), 707–716.

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., Luo, Y., 2022. Multimodal machine learning in precision health: A scoping review. Npj Digit. Med. 5 (1), 171.

Kumar, A., Raghunathan, A., Jones, R., Ma, T., Liang, P., 2022. Fine-tuning can distort pretrained features and underperform out-of-distribution. arXiv preprint arXiv:2202.10054.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Adv. Neural Inf. Process. Syst. 30.

Lavigne, H., D'ortenzio, F., Ribera D'Alcalà, M., Claustre, H., Sauzède, R., Gacic, M., 2015. On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea: a basin-scale and seasonal approach. Biogeosciences 12 (16), 5021–5039.

Law, K., Stuart, A., Zygalakis, K., 2015. Data Assimilation. 214, Springer, Cham, Switzerland, p. 52.

Lazzari, P., Álvarez, E., Terzić, E., Cossarini, G., Chernov, I., D'Ortenzio, F., Organelli, E., 2021. CDOM spatiotemporal variability in the Mediterranean Sea: a modelling study. J. Mar. Sci. Eng. 9 (2), 176.

Lazzari, P., Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S., Crise, A., 2012. Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach. Biogeosciences 9 (1), 217–233.

Lazzari, P., Solidoro, C., Salon, S., Bolzon, G., 2016. Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach. Deep. Sea Res. Part I: Ocean. Res. Pap. 108, 39–52.

LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R., 2002. Efficient backprop. In: Neural Networks: Tricks of the Trade. Springer, pp. 9–50.

Lee, D., Baldick, R., 2013. Short-term wind power ensemble prediction based on Gaussian processes and neural networks. IEEE Trans. Smart Grid 5 (1), 501–510.

Ma, Y., Zhang, D., Zhang, Y., Zhao, G., Xie, Y., Jiang, H., 2023. Advancements and challenges in deep learning-driven marine data assimilation: A comprehensive review. Comput. Res. Prog. Appl. Sci. Eng 9, 1–17.

Miloslavich, P., Seeyave, S., Muller-Karger, F., Bax, N., Ali, E., Delgado, C., Evers-King, H., Loveday, B., Lutz, V., Newton, J., et al., 2019. Challenges for global ocean observation: the need for increased human capacity. J. Oper. Ocean. 12 (sup2), S137–S156.

Na, L., Shaoyang, C., Zhenyan, C., Xing, W., Yun, X., Li, X., Yanwei, G., Tingting, W., Xuefeng, Z., Siqi, L., 2022. Long-term prediction of sea surface chlorophyll-a concentration based on the combination of spatio-temporal features. Water Res. 211, 118040.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorobot. 7, 21.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., et al., 2011. Multimodal deep learning. In: ICML. Vol. 11, pp. 689–696.

Panwar, H., Gupta, P., Siddiqui, M.K., Morales-Menendez, R., Bhardwaj, P., Sharma, S., Sarker, I.H., 2020. AquaVision: Automating the detection of waste in water bodies using deep transfer learning. Case Stud. Chem. Environ. Eng. 2, 100026.

Pearce, T., Zaki, M., Brintrup, A., Anastassacos, N., Neely, A., 2018. Uncertainty in neural networks: Bayesian ensembling. Stat 1050, 12.

Pietropolli, G., Carolina, A., Cossarini, G., Manzoni, L., 2025. GLOBIO: Bridging global and local scales for biogeochemical profiles prediction. In: EGU General Assembly Conference Abstracts. pp. EGU25–3779.

Pietropolli, G., Cossarini, G., Manzoni, L., 2022. GANs for integration of deterministic model and observations in marine ecosystem. In: EPIA Conference on Artificial Intelligence. Springer, pp. 452–463.

Pietropolli, G., Manzoni, L., Cossarini, G., 2023. Multivariate relationship in big data collection of ocean observing system. Appl. Sci. 13 (9), 5634.

Pietropolli, G., Manzoni, L., Cossarini, G., 2024. PPCon 1.0: Biogeochemical-Argo profile prediction with 1D convolutional networks. Geosci. Model. Dev. 17 (20), 7347–7364.

Pintelas, E., Pintelas, P., 2022. A 3D-CAE-CNN model for Deep Representation Learning of 3D images. Eng. Appl. Artif. Intell. 113, 104978.

Pistellato, M., Bergamasco, F., Torsello, A., Barbariol, F., Yoo, J., Jeong, J.-Y., Benetazzo, A., 2021. A physics-driven CNN model for real-time sea waves 3D reconstruction. Remote. Sens. 13 (18), 3780.

Radenović, F., Tolias, G., Chum, O., 2018. Fine-tuning CNN image retrieval with no human annotation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (7), 1655–1668.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7).

Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., Martinez, E., 2023. A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers. Front. Mar. Sci. 10, 1077623.

Russell, J., Sarmiento, J., Cullen, H., Hotinski, R., Johnson, K., Riser, S., Talley, L., 2014. The southern ocean carbon and climate observations and modeling program (soccom). Ocean. Carbon Biogeochem. Newsl 7, 1–5.

Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., Siddiqui, T., Mellit, A., 2023. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability 15 (7), 5930.

Salem, T.S., Langseth, H., Ramampiaro, H., 2020. Prediction intervals: Split normal mixture from quality-driven deep ensembles. In: Conference on Uncertainty in Artificial Intelligence. PMLR, pp. 1179–1187.

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C., Crise, A., 2019. Novel metrics based on biogeochemical argo data to improve the model uncertainty evaluation of the CMEMS mediterranean marine ecosystem forecasts. Ocean. Sci. 15 (4), 997–1022. http://dx.doi.org/10.5194/os-15-997-2019, URL https://os.copernicus.org/articles/15/997/2019/.

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. Sens. 12 (24), 4123.

Sammartino, M., Della Cioppa, L., Colella, S., Nardelli, B.B., 2025. Physically-informed Deep Neural Network for the joint prediction of chlorophyll-a and hydrographic 3D structure of the Mediterranean Sea. Environ. Model. Softw. 106660.

Santos, C.F.G.D., Papa, J.P., 2022. Avoiding overfitting: A survey on regularization methods for convolutional neural networks. ACM Comput. Surv. (Csur) 54 (10s), 1–25.

Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2018. How does batch normalization help optimization? Adv. Neural Inf. Process. Syst. 31.

Schiller, A., Mourre, B., Drillet, Y., Brassington, G., 2018. An overview of operational oceanography. New Front. Oper. Ocean. 1–26.

Schupbach, J., Sheppard, J.W., Forrester, T., 2020. Quantifying uncertainty in neural network ensembles using u-statistics. In: 2020 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. D: Nonlinear Phenom. 404, 132306.

Shrestha, A., Mahmood, A., 2019. Review of deep learning algorithms and architectures. IEEE Access 7, 53040–53065.

Šíma, J., Orponen, P., 2003. General-purpose computation with neural networks: A survey of complexity theoretic results. Neural Comput. 15 (12), 2727–2778.

Skákala, J., Awty-Carroll, K., Menon, P.P., Wang, K., Lessin, G., 2023. Future digital twins: emulating a highly complex marine biogeochemical model with machine learning to predict hypoxia. Front. Mar. Sci. 10, 1058837.

Skakala, J., Ford, D., Bruggeman, J., Hull, T., Kaiser, J., King, R.R., Loveday, B., Palmer, M.R., Smyth, T., Williams, C.A., et al., 2021. Towards a multi-platform assimilative system for North Sea biogeochemistry. J. Geophys. Res.: Ocean. 126 (4), e2020JC016649.

Stahlschmidt, S.R., Ulfenborg, B., Synnergren, J., 2022. Multimodal deep learning for biomedical data fusion: a review. Brief. Bioinform. 23 (2), bbab569.

Stow, C.A., Jolliff, J., McGillicuddy Jr., D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76 (1–2), 4–15.

Teruzzi, A., Bolzon, G., Feudale, L., Cossarini, G., 2021. Deep chlorophyll maximum and nutricline in the Mediterranean Sea: emerging properties from a multi-platform assimilated biogeochemical model experiment. Biogeosciences 18 (23), 6147–6166. http://dx.doi.org/10.5194/bg-18-6147-2021, URL https://bg.copernicus.org/articles/18/6147/2021/.

Teruzzi, A., Dobricic, S., Solidoro, C., Cossarini, G., 2014. A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties. J. Geophys. Res.: Ocean. 119 (1), 200–217.

Tonelli, T., Pietropolli, G., Cossarini, G., Manzoni, L., 2025. Convolutional neural networks for chlorophyll prediction in the Mediterranean Sea. Technical Report, Copernicus Meetings.

Tripathy, R.K., Bilionis, I., 2018. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. J. Comput. Phys. 375, 565–588.

Twomey, J.M., Smith, A.E., et al., 1997. Validation and verification. Artif. Neural Netw. Civ. Eng.: Fundam. Appl. 44–64.

Verdy, A., Mazloff, M.R., 2017. A data assimilating model for estimating S outhern O cean biogeochemistry. J. Geophys. Res.: Ocean. 122 (9), 6968–6988.

Vidussi, F., Marty, J.-C., Chiavérini, J., 2000. Phytoplankton pigment variations during the transition from spring bloom to oligotrophy in the northwestern Mediterranean Sea. Deep. Sea Res. Part I: Ocean. Res. Pap. 47 (3), 423–445.

Volpe, G., Nardelli, B.B., Colella, S., Pisano, A., Santoleri, R., 2018. An operational interpolated ocean colour product in the Mediterranean sea. New Front. Oper. Ocean. 227–244.

Walach, E., Wolf, L., 2016. Learning to count with cnn boosting. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, pp. 660–676.

Wang, H., Hu, S., Li, X., 2023. An interpretable deep learning ENSO forecasting model. Ocean-Land-Atmosphere Res. 2, 0012.

Wang, H., Li, X., 2023. DeepBlue: Advanced convolutional neural network applications for ocean remote sensing. IEEE Geosci. Remote. Sens. Mag. 12 (1), 138–161.

Wang, X., Wang, R., Hu, N., Wang, P., Huo, P., Wang, G., Wang, H., Wang, S., Zhu, J., Xu, J., et al., 2024. Xihe: A data-driven model for global ocean eddy-resolving forecasting. arXiv preprint arXiv:2402.02995.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3, 1–40.

White, F.E., et al., 1991. Data fusion lexicon. Jt. Dir. Lab. Tech. Panel C 3, 19.

Yan, X., Du, J., Wang, L., Liang, Y., Hu, J., Wang, B., 2024. The synergistic role of deep learning and neural architecture search in advancing artificial intelligence. In: 2024 International Conference on Electronics and Devices, Computational Science. ICEDCS, IEEE, pp. 452–456.

Yatheendradas, S., Kumar, S., 2022. A novel machine learning–based gap-filling of fine-resolution remotely sensed snow cover fraction data by combining downscaling and regression. J. Hydrometeorol. 23 (5), 637–658.

Zanna, L., Bolton, T., 2021. Deep learning of unresolved turbulent ocean processes in climate models. Deep. Learn. the Earth Sci.: A Compr. Approach To Remote. Sens. Clim. Sci. Geosci. 298–306.

Zhang, Y., Sevault, F., Pennel, R., Baklouti, M., 2025. Analysis of the impact of vertical variation and temporal frequency of the chlorophyll forcing field on modelled temperature in the Mediterranean Sea and potential implications for regional climate projections. Ocean. Model. 194, 102490.

Zhang, Q., Sun, W., Guo, H., Dong, C., Zheng, H., 2024. A transfer learning-enhanced generative adversarial network for downscaling sea surface height through heterogeneous data fusion. Remote. Sens. 16 (5), 763.

Zhang, Y., Thorburn, P.J., 2022. Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. Future Gener. Comput. Syst. 128, 63–72.

Zhang, X., Wang, H., Wang, S., Liu, Y., Yu, W., Wang, J., Xu, Q., Li, X., 2022. Oceanic internal wave amplitude retrieval from satellite images based on a data-driven transfer learning model. Remote Sens. Environ. 272, 112940.

Zhao, H., Gallo, O., Frosio, I., Kautz, J., 2016. Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging 3 (1), 47–57.

Zhao, X., Gong, X., Gong, X., Liu, J., Wang, G., Wang, L., Guo, X., Gao, H., 2024. Evolution of 3-D chlorophyll in the northwestern Pacific ocean using a Gaussian-activation deep neural network model. Front. Mar. Sci. 11, 1378488.

Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., Zhong, G., 2024. Applications of deep learning in physical oceanography: a comprehensive review. Front. Mar. Sci. 11, 1396322.

Zheng, Y., Li, C., Zhou, W., Xu, Z., Zhang, X., Cao, W., Yang, Z., Liu, C., 2024. A novel method to estimate the 3D chlorophyll a distribution in the South China Sea surface waters using hydrometeorological parameters. Sci. Rep. 14 (1), 25516.

Zhou, Y., Wang, H., Xu, F., Jin, Y.-Q., 2016. Polarimetric SAR image classification using deep convolutional neural networks. IEEE Geosci. Remote. Sens. Lett. 13 (12), 1935–1939.

Zhou, S., Xie, W., Lu, Y., Wang, Y., Zhou, Y., Hui, N., Dong, C., 2021. ConvLSTM-based wave forecasts in the South and East China seas. Front. Mar. Sci. 8, 680079.