



A Deep Learning Framework for Chlorophyll Prediction in Large Marine Ecosystems: Benchmarking with a Dynamic Model and Implications for Fish Catch Forecasts

5 Ji-Sook Park¹, Jong-Yeon Park^{1,2*}, Yoo-Geun Ham^{3*}, Jeong-Hwan Kim⁴, and WooJin Jeon¹

¹Department of Climate, Environment, and Energy, Jeonbuk National University, Jeonju, 54896, South Korea

²Department of Earth and Environmental Sciences & Earth Environmental System Research Center, Jeonju, 54896, South Korea

³Department of Environmental Management, Seoul National University, Seoul, 08826, South Korea

10 ⁴Center for Climate and Carbon Cycle Research, Korea Institute of Science and Technology, Seoul, 02792, South Korea

Correspondence to: Jong-Yeon Park (jongyeon.park@jbnu.ac.kr), Yoo-Geun Ham (teru81@gmail.com)

Abstract. Anticipating marine ecosystem changes is critical for enabling communities to adapt to climate fluctuations and for predicting future climate by considering interactions between Earth's physical and biogeochemical fields. Earth System
15 Models (ESMs) simulate Earth's multi-facet features, but their predictive capabilities remain limited due to sparse biogeochemical observations and structural uncertainties in marine biogeochemical models. Here, we develop a deep learning-based prediction system to forecast surface chlorophyll concentrations across all Large Marine Ecosystems (LMEs). Trained on multi-decadal simulations from various climate models and a coupled physical-biogeochemical reanalysis from a data assimilative ESM run, the system demonstrates skillful chlorophyll predictions comparable to ESM-
20 based dynamic forecasts. The prediction skill arises from physical-biogeochemical coupling processes triggered by large-scale climate variability, consistent with the mechanisms previously identified in dynamical forecasts. Furthermore, predicted chlorophyll anomalies are significantly linked to interannual variability in fish catch in several LMEs, demonstrating the promise of data-driven biogeochemical forecasting to support adaptive, climate-informed marine resource management.

25 1 Introduction

Marine ecosystem plays a pivotal role in regulating Earth's climate system, particularly through the cycling of carbon and other greenhouse gases at the ocean-atmosphere boundary (Volk and Hoffert, 1985; Falkowski et al., 2000). Phytoplankton, a central component of the marine ecosystem, drives the biological carbon pump via photosynthesis (Falkowski et al., 1998; Field et al., 1998) and also modulates the physical properties of the ocean surface, such as surface albedo and the vertical
30 distribution of solar shortwave radiation, thereby influencing upper ocean temperature (Sweeney et al., 2005; Park et al.,



2018a). These biogeochemical and biogeophysical feedbacks can affect large-scale climate variability and long-term global warming patterns across multiple timescales. Understanding and predicting marine biogeochemical variability is therefore critical for advancing climate predictions based on bio-climate interactions and supporting the sustainable management of marine ecosystems (Bonan and Doney, 2018; Capotondi et al., 2019; Siegel et al., 2023).

35 However, translating this understanding into actionable prediction remains challenging. Earth System Models (ESMs), which integrate biogeochemical processes within physical climate frameworks, serve as primary tools for simulating coupled ocean–climate interactions. While these models have demonstrated skillful forecasts of oceanic physical variables, including sea surface temperature, upper-ocean heat content, and large-scale variability, on seasonal to decadal timescales (Smith et al., 2020; Balmaseda et al., 2024), their performance for biogeochemical variables remains limited. Recent advances have shown
 40 skillful chlorophyll forecasts in many regions (Park et al., 2019), yet fundamental challenges persist. These include limited three-dimensional observations of biogeochemical fields, with satellite-derived chlorophyll-a records extending only since the late 1990s (Henson et al., 2010; Henson et al., 2016), which restricts model initialization and contributes to substantial uncertainties in biogeochemical parameterizations (Séférian et al., 2020; Fennel et al., 2022). Additionally, large inter-model discrepancies remain, especially in regions with poor observation coverage, where insufficient constraints on model
 45 parameters amplify uncertainty (Mignot et al., 2023; Kwiatkowski et al., 2020). Moreover, the substantial computational costs required for extensive ensemble experiments needed for rigorous skill assessment further constrain the operational utility of current ESM approaches (Balaji et al., 2022). These limitations have highlighted the need for alternative methodologies that can provide skillful biogeochemical forecasts with greater computational efficiency.

Deep learning has emerged as a promising alternative for predicting marine biogeochemical variability in the presence of
 50 observational gaps and structural uncertainties in marine biogeochemical models. These data-driven models can learn complex, nonlinear relationships from limited or noisy inputs, making them well-suited for data-sparse components of the Earth system (Reichstein et al., 2019). Deep learning approaches have shown skills in forecasting physical ocean variables—such as sea surface temperature, sea level, and major climate modes including El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD)—across various timescales (Biswas and Sinha, 2021; Xiao et al., 2019; Song et al., 2020; Immas
 55 et al., 2021; Ham et al., 2019). Building on this, recent studies have applied deep learning to biogeochemical domains, including historical chlorophyll-a reconstruction (Roussillon et al., 2023), phytoplankton biomass estimation (Yu et al., 2020), satellite data gap-filling (Hong et al., 2023), and biogeochemical forecasting applications in regional marine systems (Cen et al., 2022; Yao et al., 2023). Despite this progress, existing efforts often lack global spatial scope, suffer from limited interpretability due to their "black box" nature, show insufficient integration with physics-based models.

60 To address these limitations, we developed a global-scale forecasting framework based on a convolutional neural network (CNN) to predict surface chlorophyll across all Large Marine Ecosystems (LMEs), coastal domains that support the majority of the world's marine fish catch (Fig. 1). The model ingests three consecutive months of global sea surface temperature and chlorophyll anomalies and produces monthly or annual chlorophyll forecasts at the LME scale. The training dataset consists of multi-decadal simulations from the Coupled Model Intercomparison Project phase 6 (CMIP6) multi-model ensemble



(Eyring et al., 2016) and physical–biogeochemical reanalysis fields from an ESM-based data assimilation system developed at the NOAA Geophysical Fluid Dynamics Laboratory (GFDL). The assimilation system incorporates three-dimensional physical observations into a coupled physical–biogeochemical model to reconstruct historical ocean states (Park et al., 2018b). Model predictions are validated against satellite-derived chlorophyll from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) (Lundberg et al., 2020; Lundberg et al., 2018), and compared with ESM-based dynamical biogeochemical forecasts.

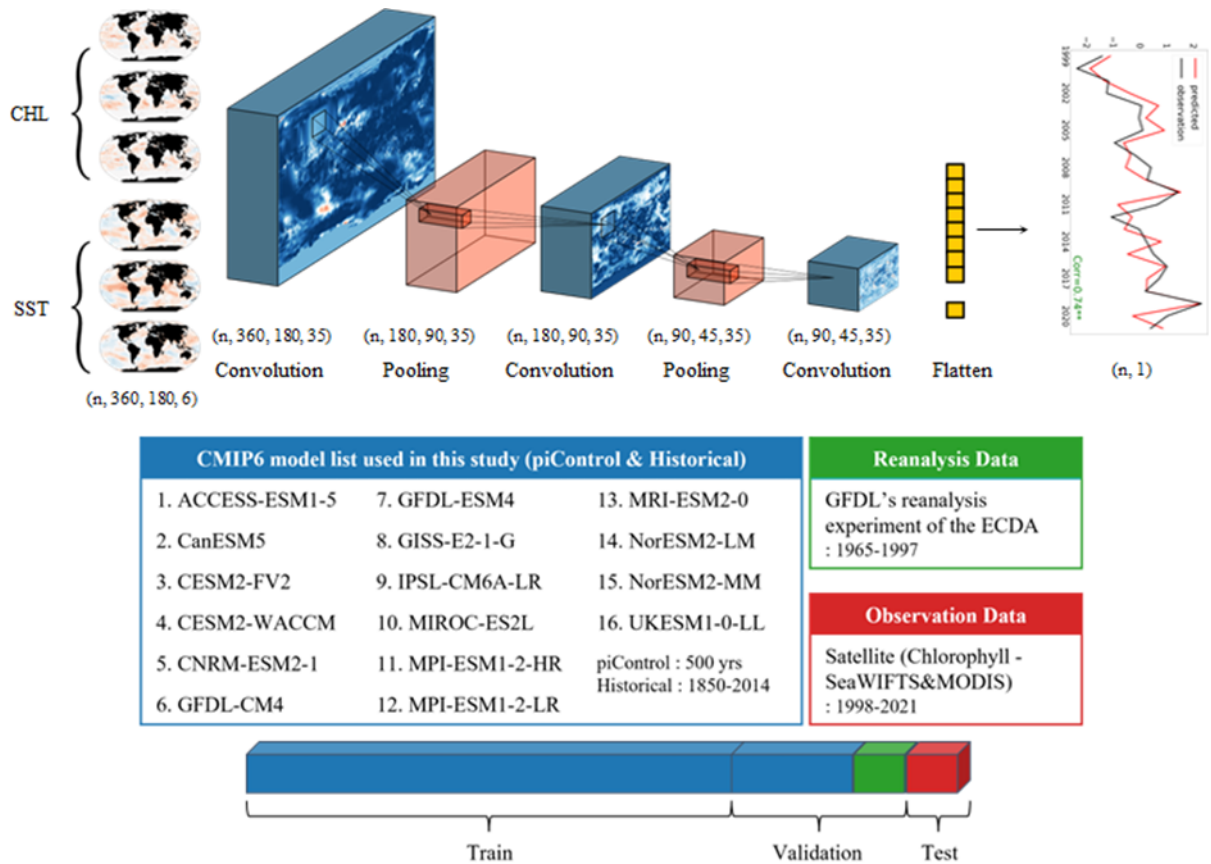


Figure 1: Deep learning model structure. The adapted CNN model comprises three convolutional layers (blue), two max-pooling (MP) layers (red), and one fully connected (FC) layer (yellow). Input data include sea surface temperature (SST) and chlorophyll anomalies for three consecutive months (e.g., November–January), represented as six channels. The model predicts either monthly or annual mean chlorophyll anomalies for each Large Marine Ecosystem (LME). Training data comprise historical and piControl simulations from 16 CMIP6 models, along with physical–biogeochemical reanalysis (1965–1997). Model validation was performed using satellite-based chlorophyll observations from SeaWiFS and MODIS (1998–2021).



2 Methods

2.1 Deep learning model and forecast experiment design

80 The convolutional neural network (CNN) model used in this study was adapted from prior work on spatiotemporal prediction (Ham et al., 2019). It consists of an input layer followed by three convolutional layers, two max-pooling layers, one fully connected layer, and an output layer. The network incorporates 35 convolutional filters per layer, uses Gaussian Error Linear Unit (GELU) activation functions (Hendrycks and Gimpel, 2016), and is trained using the Mean Absolute Error (MAE) loss function.

85 Input data for the deep learning model consist of three consecutive months of SST and chlorophyll anomalies (gridded at $1^\circ \times 1^\circ$ resolution), enabling the prediction of seasonal or annual mean surface chlorophyll anomalies for each Large Marine Ecosystem (LME). The model was trained using both historical and piControl simulations from CMIP6 and validated using GFDL-ECDA reanalysis (Park et al., 2018b). Satellite-derived chlorophyll served exclusively for final testing. Predictions were initialized seasonally to forecast annual chlorophyll anomalies with lead times up to 24 months. A series of sensitivity
 90 experiments was performed to identify the optimal model configuration.

Prediction performance for seasonal or annual chlorophyll anomalies was evaluated using the anomaly correlation coefficient (ACC) between predicted and satellite chlorophyll anomalies at the LME scale. Statistical significance of correlations was assessed following a method adapted from previous work (Bretherton et al., 1999), accounting for temporal autocorrelation through effective degrees of freedom.

95 2.2 Data sources and preprocessing

Satellite monthly surface chlorophyll-a concentrations were obtained from the SeaWiFS and MODIS ocean color sensors (Esaias et al., 1998; Mcclain et al., 1998), and sea surface temperature (SST) data were from NOAA's optimally interpolated SST version 2 (OISSTv2) dataset based on the Advanced Very High Resolution Radiometer (AVHRR) (Reynolds et al., 2007). The original chlorophyll and SST data were provided at daily resolution with fine spatial scales (0.25 degrees for SST
 100 and $9 \text{ km} \times 9 \text{ km}$ for chlorophyll). For consistency and computational efficiency in deep learning applications, all observational data spanning 1998 to 2021 were resampled to monthly means and interpolated onto a $1^\circ \times 1^\circ$ regular global grid. Due to cloud cover and persistent polar night, the ocean color datasets contained missing values. A standard zero-filling strategy was applied, where pixels with missing values were identified, masked, and filled with zeros. The same preprocessing scheme, including temporal aggregation and spatial resampling, was uniformly applied to both chlorophyll
 105 and SST datasets used for training, validation, and testing.

Reanalysis data used for validation and sensitivity testing were obtained from the NOAA Geophysical Fluid Dynamics Laboratory's Ensemble Coupled Data Assimilation (GFDL-ECDA) system, integrated with the COBALT biogeochemical model (Park et al., 2018b). This system assimilates observed physical variables into a coupled physical-biogeochemical

framework while excluding direct assimilation of biogeochemical variables to avoid spurious vertical velocity artifacts near the equator. Data from 1965 to 1997 were used for validation, while the 1998–2017 period supported sensitivity analysis. Long-term simulated chlorophyll and SST data were drawn from historical and preindustrial control (piControl) runs of 16 models from the Coupled Model Intercomparison Project Phase 6 (CMIP6). While simulated chlorophyll underwent the same gap-filling treatment, the SST fields were not gap-filled but were resampled to the same spatiotemporal grid for consistency. Given variability in the length of piControl simulations among models, only the most recent 500 years were used when available; for models with shorter records, the entire simulation period was included. Historical simulations covering 1850–2014 were also used. CMIP6 simulations were employed exclusively for training and validation purposes.

2.3 SHAP Analysis

To interpret the model's predictions and identify dominant spatial drivers, we applied SHapley Additive exPlanations (SHAP) (Lundberg et al., 2020; Lundberg et al., 2018). SHAP provides feature-level attributions by estimating the marginal contribution of each input (grid cell) to the final model output. For each prediction, the SHAP decomposition follows Eq. (1):

$$y = \bar{y} + \sum_i \varphi_i, \quad (1)$$

where y is the prediction, $\bar{y}$ is the mean prediction, and φ_i is the SHAP value (i.e., the contribution) of feature i , which in this case corresponds to a specific grid point in the input map.

Because the target variable is chlorophyll concentration anomalies, which can be positive or negative, we analyze the absolute SHAP values to interpret how each grid point contributes to pushing the anomalies in either the positive or negative direction. Large absolute SHAP values indicate that the input conditions at a particular grid point have a strong influence on the predicted chlorophyll anomaly for the region of interest. SHAP values are calculated by estimating the marginal contribution of each grid point across all possible permutations of the input map. This is done by comparing the model's predictions with and without the grid point of interest, while considering all possible subsets of the other grid points.

2.4 Fish catch prediction

We utilized annual reported fish catch data from the Sea Around Us project (Pauly and Zeller, 2016), which compiles species-resolved annual harvests by LME. Total annual catches per species were calculated, and ambiguous or non-specific entries were excluded. To identify the ten most harvested fish species in each LME, the top ten species were selected by cumulative catch volume. Catch anomalies were computed as normalized values by subtracting the mean and dividing by the standard deviation for each species.

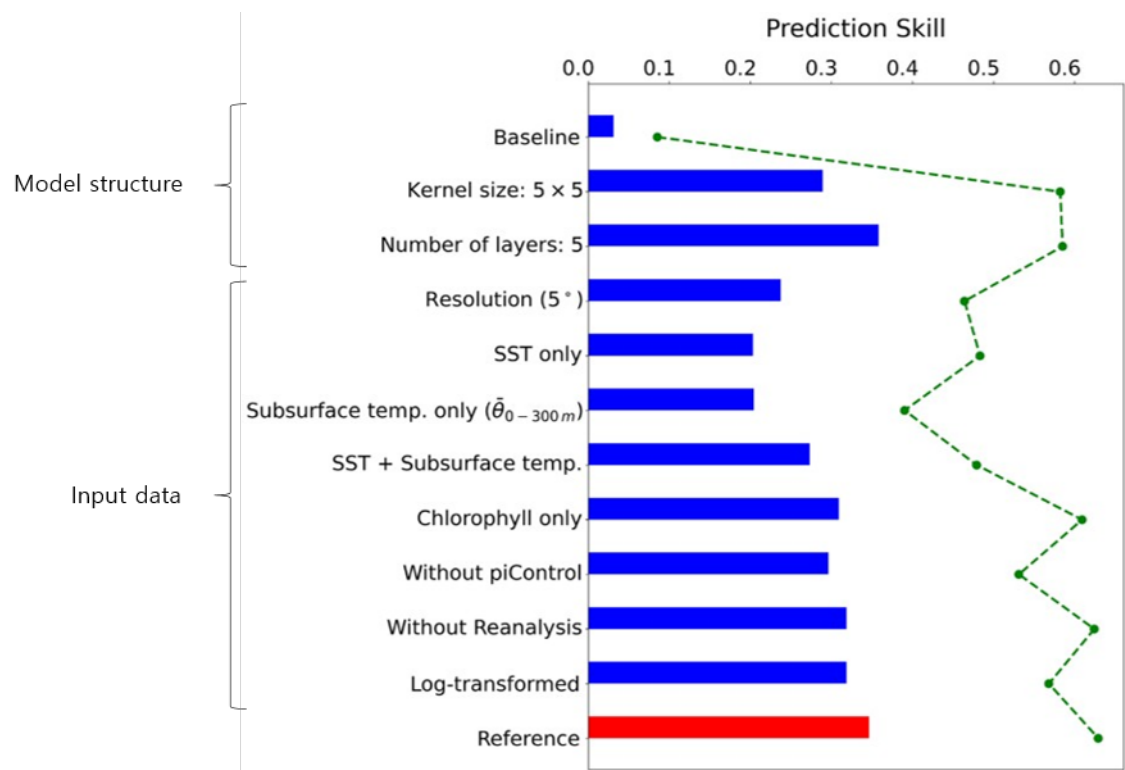
To examine the linkage between chlorophyll predictions and fisheries, we employed simple linear regression models. Specifically, we used chlorophyll anomaly forecasts initialized in November–January (NDJ) to predict annual fish catch anomalies in selected LMEs. Statistical significance of correlation coefficients between predicted and reported fish catch was



140 assessed using effective degrees of freedom to account for temporal autocorrelation (Bretherton et al., 1999), similar to the statistical test for chlorophyll prediction. The linear regression analysis was performed with statistical significance evaluated at $p < 0.05$ and $p < 0.10$ levels.

3 Results

3.1 Sensitivity experiments of model configuration



145 **Figure 2: Sensitivity test of model configuration.** Bars indicate the average correlation skill across 16 selected regions for each model variation. All configurations are derived from the reference model (red bar), which exhibited the highest overall predictive performance. In each sensitivity experiment (blue bars), a single component of the reference model was modified, either a structural aspect (e.g., kernel size, number of layers) or input data configuration (e.g., resolution, predictor variables, log transformation). The baseline model, shown at the top, shares the same architecture as the reference model but differs in training settings, including weight initialization and loss function. The green dashed line shows the average skill across regions where prediction was statistically significant ($p < 0.10$) in at least one configuration.

150

A systematic evaluation of model configurations was performed to assess how architectural and data choices affect chlorophyll prediction skill across global marine ecosystems (Fig. 2). For computational efficiency, the sensitivity experiments were conducted by selecting 16 representative LMEs, and the average prediction skill across these regions was used for comparison. The reference model, optimized through these sensitivity experiments, showed high skill in forecasting

155



annual chlorophyll anomalies across LMEs during 1998-2021. This optimized configuration employed three convolutional layers with 3×3 kernels, GELU activation functions, and mean absolute error loss. Compared with a baseline model using standard settings, such as ReLU activation functions and mean squared error loss, the reference model substantially outperformed this baseline, reflecting the importance of task-specific optimization for marine ecosystem prediction.

To assess the sensitivity of model performance to individual components of the reference configuration, we modified one element at a time while keeping all other settings constant. For example, replacing the 3×3 convolutional kernels in the reference model with broader 5×5 kernels reduced prediction skill relative to the reference configuration, despite the increase in trainable parameters. This suggests that the smaller kernel size is more effective at capturing local structure while still integrating small-scale biogeochemical structures. This result aligns with established principles in CNN architecture design, where smaller kernels can effectively capture local features without compromising model performance.

In addition to architectural considerations, model performance was highly sensitive to input data configuration and preprocessing. High-resolution (1°) input data produced markedly higher predictive skill than coarser (5°) fields, reflecting the importance of resolving mesoscale variability that drives chlorophyll dynamics (Keerthi et al., 2022). Predictor selection for the input data proved equally critical: models trained with surface chlorophyll anomalies as input substantially outperformed those using only physical variables, such as sea surface temperature (SST) or subsurface potential temperature. This suggests that surface chlorophyll captures nonlinear, ecologically relevant signals potentially linked to subsurface processes or serving as early indicators of climate-driven ecosystem shifts as shown in previous studies (Park et al., 2018a).

The inclusion of additional input datasets generally improved the model's prediction skill. Incorporating CMIP6 piControl simulations, designed to represent long-term natural variability in the absence of anthropogenic forcing, enhanced the model performance, despite the temporal gap with present-day climate conditions. Similarly, inclusion of the ESM-based reanalysis product, which provides physically consistent and observationally constrained ocean states, also improved chlorophyll prediction skill by enabling the model to learn from longer historical biogeochemical variability informed by climate-scale physical dynamics.

Finally, the impact of applying a log transformation to chlorophyll data on the prediction skill was also tested. While log transformation is often used to normalize skewed chlorophyll distributions, our results indicate that retaining the original scale shows better prediction skills. This advantage suggests that preserving the raw variability better captures the full range of biogeochemical signals relevant for ecosystem prediction, as log-scaling dampens high chlorophyll variability, leading the deep learning model to underestimate chlorophyll, particularly in coastal regions with high concentrations (Cen et al., 2022). Overall findings here informed the development of an optimized model configuration that combines an efficient model architecture, high-resolution inputs, ecologically meaningful predictors, and physically consistent long-term training data. Rather than representing ad hoc tuning, this configuration reflects deliberate design choices informed by empirical performance and domain knowledge. The reference model, which achieved the highest prediction skill across the LMEs, serves as the foundation for all subsequent analyses, including model validation against satellite data, investigation of the mechanisms driving skillful predictions, and applications to fish catch forecasting.

3.2 LME-scale chlorophyll prediction

The optimized reference model derived from the sensitivity experiments was applied across all global LMEs to evaluate its skill in forecasting monthly to annual chlorophyll anomalies. Annual forecasts were initialized in early winter (November to January) of the preceding year and tested with different combinations of physical and biogeochemical inputs, that is, SST only, chlorophyll only, and both SST and chlorophyll, to assess the contribution of each predictor.

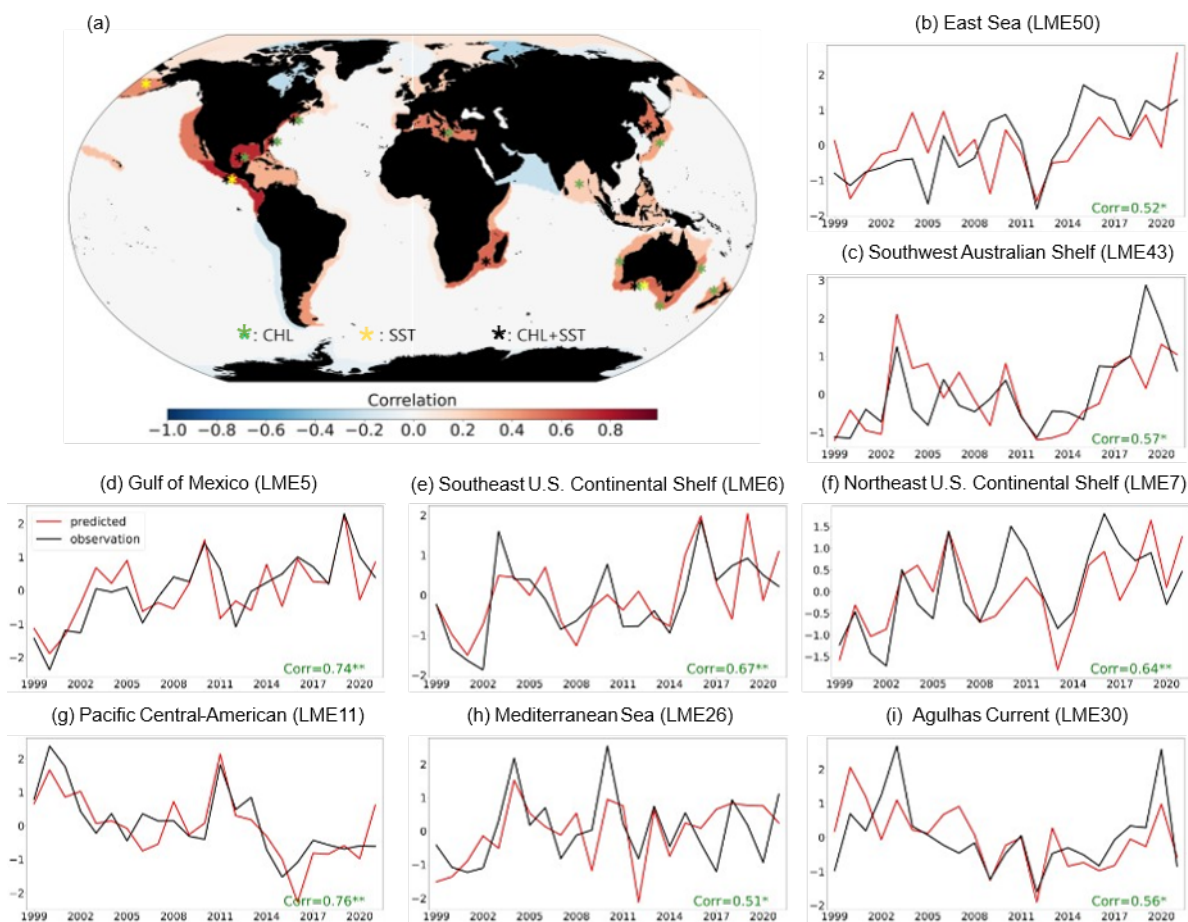


Figure 3: Chlorophyll prediction skill in Large Marine ecosystem (LME). a Correlation coefficients between satellite-derived annual chlorophyll and predicted annual chlorophyll. Significant correlations ($P < 0.1$) are denoted by asterisks. Shading represents the prediction skill using both chlorophyll (CHL) and sea surface temperature (SST). Green asterisks indicate the region with a significant correlation using only chlorophyll, yellow asterisks using only SST, while black asterisks using both variables (forecast lead time = 1 year). b-i Time series data for LMEs with significant prediction skill when both chlorophyll and SST inputs are used (corresponding to black asterisks in panel (a)). Panels show normalized annual mean chlorophyll anomalies from satellite (black) and model predictions (red). Correlation values are indicated with significance levels (*: $P < 0.1$, **: $P < 0.05$).

The model demonstrated skillful annual mean chlorophyll predictions in many LMEs (Fig. 3). Statistically significant skill indicated by asterisks in Fig. 3a was observed in regions along the Central-North American coast, the western Indian Ocean,



and the Australian coast. Among the regions showing successful predictions, eight LMEs using both SST and chlorophyll inputs were selected to compare the time series of predicted and satellite chlorophyll. The predicted annual chlorophyll successfully captured both interannual and long-term variability, closely following satellite-derived chlorophyll. One notable point here is that when the model was trained using only SST as input, significant predictive skill was achieved in only three LMEs. In contrast, including surface chlorophyll anomalies, either alone or as an additional predictor, substantially increased the number of LMEs where the model achieved high prediction skill. This result indicates the benefit of incorporating biogeochemical variables into data-driven models alongside physical predictors in improving chlorophyll forecasts.

Chlorophyll prediction skill was further evaluated at shorter time scales. We examined monthly forecasts by selecting two representative systems from Pacific and Indian Oceans, exhibiting significant annual mean chlorophyll prediction skill: the Pacific Central-American Coastal (LME 11) and the Agulhas Current (LME 30). For each LME, we conducted monthly forecasts from 1998 to 2021, with each forecast extending up to 24 months. Each forecast used three consecutive months of input data and employed five ensemble members per initialization. Model predictions were compared to satellite-derived chlorophyll anomalies after applying a 3-month moving average to facilitate skill assessment at seasonal scales.

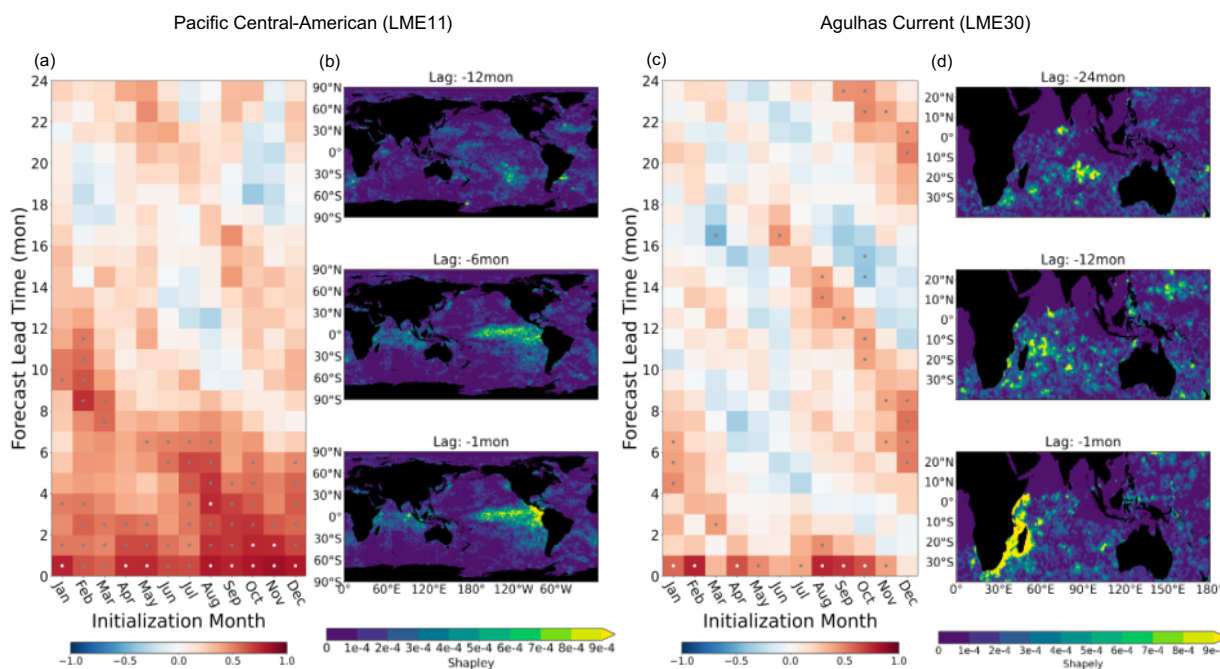


Figure 4: Monthly prediction and mechanism underlying chlorophyll prediction skill. a,c Anomaly correlation skill as a function of forecast initialization month (x-axis) and lead time (y-axis). White dots indicate significant skill at $P < 0.05$, while grey dots indicate $P < 0.10$. Correlation coefficients are based on 3-month running mean anomalies. b,d Temporal evolution of spatial maps of absolute Shapley values. Panels (a) and (b) correspond LME 11 (Pacific Central-American Coastal), while panels (c) and (d) correspond to LME 30 (Agulhas Current).



In the Pacific Central-American region, the model exhibits seasonally varying forecast skill, with statistically significant correlations extending up to 12-month lead times for forecasts initialized during boreal winter (Fig. 4a). Prediction skill for chlorophyll is enhanced during boreal fall and winter, when large-scale climate variability such as ENSO is more predictable, but diminished during boreal spring and early summer, coinciding with the well-documented “spring predictability barrier” of ENSO. These patterns suggest that the model leverages ENSO-related signals to enhance chlorophyll prediction in this region, consistent with previous observational and modeling studies of primary productivity in the tropical Pacific (Park et al., 2019; Pennington et al., 2006; Sasai et al., 2012).

In the Agulhas Current LME, the model exhibited a seasonally modulated pattern of forecast skill, marked by alternating bands of high and low correlation that persisted across lead times up to 24 months (Fig. 4c). This diagonal structure is particularly pronounced for austral winter initializations and resembles the winter-to-winter reemergence mechanism observed in dynamical prediction systems. In this process, wintertime anomalies are subducted beneath the mixed layer, preserved during summer stratification, and reemerge the following winter as seasonal mixing deepens the surface layer. The recurrence of this pattern in the model’s predictions indicates that it captures subsurface ocean memory in addition to surface signals.

3.3 Mechanisms underlying chlorophyll prediction skills

To examine the physical basis of the regional chlorophyll forecast skill, we applied SHapley Additive exPlanations (SHAP) to quantify the contribution of input features across lead times, focusing on the two regions where monthly forecasts were conducted. In the Pacific Central-American region, we examined boreal winter 2014–2015, which captured the early development phase of El Niño conditions, as documented by satellite chlorophyll observations. Attribution maps from the models initialized from this period reveal coherent patterns at 1-, 6-, and 12-month horizons, aligning with the canonical progression of ENSO-related anomalies, including the emergence and eastward propagation of SST signals along the equatorial Pacific (Fig. 4b). While SHAP does not infer causality, the spatial alignment between feature importance and known ENSO structures shows that the deep learning model can detect climate-scale variability relevant to chlorophyll prediction.

Attribution analysis during 2000–2002 in the Agulhas region, a period of peak chlorophyll concentrations in the region, revealed westward-propagating chlorophyll anomalies originating in the eastern Indian Ocean and extending toward the western boundary (Fig. 4d). This pattern is consistent with the dynamics of upwelling Rossby waves, which have been previously identified as key contributors to long-lead chlorophyll predictability in ESM-based dynamical forecasts in this region (Jeon et al., 2022). The presence of such physically interpretable propagation features indicates that the model captures spatiotemporal dynamics embedded in the training data, beyond capturing surface-level statistical associations.

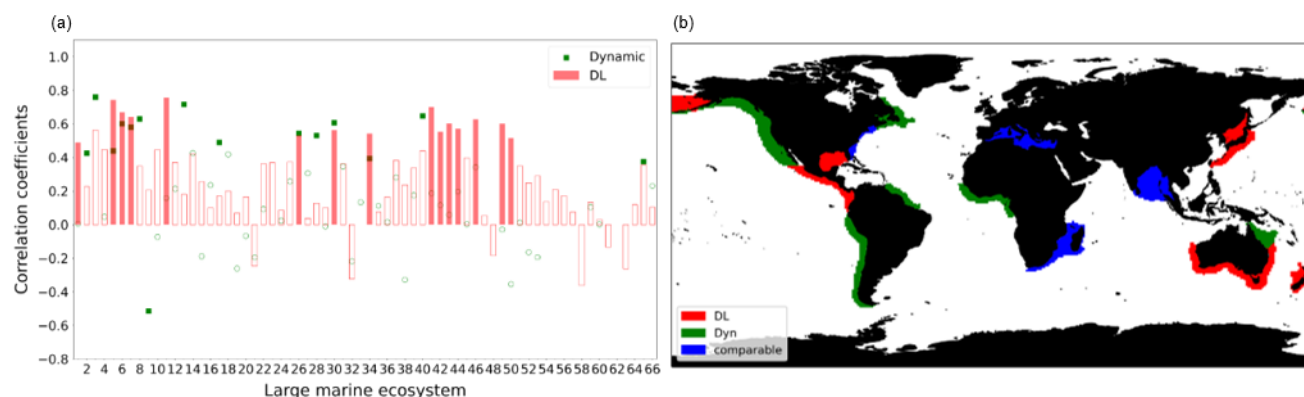
Results from the both Pacific Central-American Coastal and Agulhas Current LMEs demonstrate that the deep learning model captures statistical regularities as well as physically interpretable signals underlying chlorophyll variability. SHAP-based attribution reveals mechanisms consistent with previously-established ocean–climate processes, including ENSO



260 evolution, wintertime reemergence, and westward-propagating off-equatorial Rossby waves, suggesting that the model internalizes aspects of coupled physical–biogeochemical dynamics from the training data. These findings highlight the potential of data-driven approaches to support mechanistically informed, climate-relevant biogeochemical forecasts.

3.4 Prediction skill comparison with dynamic forecasts

We next compared the predictive performance of our deep learning model with that of a dynamical prediction system to
 265 assess relative skill. Chlorophyll prediction skill was evaluated against an ESM-based biogeochemical prediction system across global LMEs. This dynamical system, developed at the Geophysical Fluid Dynamics Laboratory (GFDL), builds on a seasonal climate prediction framework with coupled ocean-atmosphere data assimilation system and is run with a marine ecosystem model, the Carbon, Ocean Biogeochemistry and Lower Tropics (COBALT) (Zhang et al., 2007; Stock et al., 2014). The retrospective predictions were initialized on the first day of each calendar month from 1991 to 2017 and consist
 270 of 2 year-long forecasts with 12-member ensembles (see Park et al., 2019, for details).



275 **Figure 5: Comparison of prediction skills for chlorophyll anomalies in Large Marine Ecosystems (LMEs).** a Correlation coefficients between satellite-derived and predicted annual mean chlorophyll anomalies in LMEs at a 1-year lead time. The deep learning model (1998–2021; red bars) and the dynamic model (1998–2017; green scatter plots) are compared. Filled bars and marks indicate significance at the 90% level ($P < 0.10$). b Map representation of panel (a), comparing the prediction skill of the deep learning model with that of the dynamic model. Areas where one model outperforms the other are highlighted, with a correlation difference ≥ 0.2 at the 90% significance level indicating model strength. The deep learning model is shown in red, and the dynamic model in green. Areas with a correlation difference < 0.2 are depicted in blue. Non-significant or NaN regions are not displayed.

280 Prediction skills between the deep learning and dynamic models were assessed using correlation coefficients between predicted and satellite-derived annual chlorophyll anomalies at a 1-year lead time (Fig. 5a). Although skill varied regionally, both models showed a similar number of LMEs with statistically significant correlations. They also exhibited comparable performance in regions such as the Bay of Bengal, Mediterranean Sea, and Agulhas Current, where correlation values
 285 exceeded 0.6. These regions are strongly influenced by basin-scale climate modes such as ENSO and the IOD, whose surface signals are well captured by both physical and biogeochemical predictors (Fiedler, 2002; Beal and Bryden, 1999).



A categorical global map of relative performance shows that the deep learning model outperformed the dynamical system in several coastal LMEs, particularly around northern Australia and along the Pacific Central-American coast (Fig. 5b). These regions exhibit complex chlorophyll-SST relationships that likely reflect the integrated effects of multiple environmental drivers. The data-driven approach of deep learning appears well-suited to identifying predictive patterns in these surface variables without requiring explicit parameterization of underlying processes. Feature attribution analyses further support this interpretation, consistently highlighting the contributions of climate-sensitive predictors such as surface chlorophyll and SST (Amorim et al., 2021; Liu et al., 2025).

In contrast, the deep learning model showed lower prediction skill relative to the dynamical system in some regions, most notably the Pacific Eastern Boundary Upwelling Systems, including the Humboldt and California Currents. These regions are strongly influenced by wind-driven upwelling and episodic vertical nutrient fluxes, which are explicitly resolved in dynamical models with process-based parameterizations. While surface chlorophyll can partially reflect subsurface variability, especially in regions with coherent thermocline dynamics (Park et al., 2018a), such signals may be too intermittent or weakly expressed at the surface in those upwelling zones. This limits the ability of surface-based predictors to capture the timing and magnitude of upwelling-driven productivity variations, and likely contributes to the superior performance of the dynamical model in these physically dominated systems.

Overall, the results here suggest the importance of aligning model inputs with the dominant processes in each system. The deep learning model performed well in coastal LMEs characterized by complex and nonlinear dynamics, such as the northern Australian coast and the Pacific Central-American region, where dynamical models often face skill limitations due to the spring predictability barrier. These regional performance patterns reflect the different ways each modeling approach captures biogeochemical variability. Taken together, these findings suggest that deep learning and dynamical approaches offer complementary strengths. Hybrid frameworks that combine data-driven learning with physically informed constraints may help improve biogeochemical forecasts across diverse marine environments.

3.5 Fish catch prediction

The successful prediction of chlorophyll anomalies in many coastal LMEs suggests potential applications for marine resource management. To demonstrate the practical utility of deep learning based biogeochemical forecasting, we evaluated the capacity to anticipate interannual fish catch variations using predicted chlorophyll anomalies as environmental drivers. We focused on key species in selected LMEs where the deep learning model demonstrated high predictive skill. For each region, the ten most frequently caught species were identified, and linear regression was applied to assess the relationship between chlorophyll anomalies and species-specific catch anomalies.

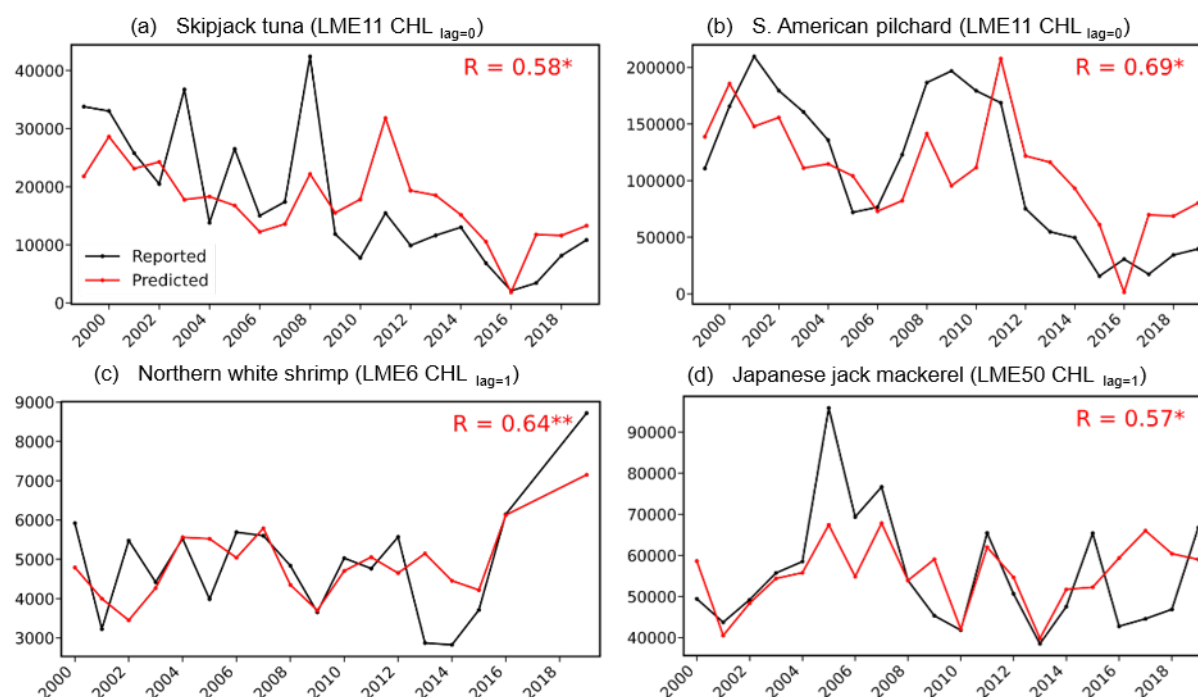


Figure 6: Prediction skill for annual fish catch of individual species in Large Marine Ecosystems (LMEs). a–d Time series show correlations between predicted and reported annual fish catches for dominant species in selected LMEs, based on chlorophyll as the primary predictor with varying time lags. (a) LME 11, lag 0: correlation coefficients for Skipjack tuna. (b) LME 11, lag 0: correlation coefficients for South American pilchard. (c) LME 6, lag 1: correlation coefficients for Northern white shrimp. (d) LME 50, lag 1: correlation coefficients for Japanese jack mackerel. Asterisks indicate significant anomaly correlation skills (* $P < 0.1$, ** $P < 0.05$). Predictions are initialized in NDJ (November(0), December(0), January(1)), with each panel representing one of the top 10 dominant species in the respective LME.

The results revealed distinct regional and species-level associations (Fig. 6). In the Pacific Central-American Coastal LME (LME 11), statistically significant contemporaneous correlations were found for skipjack tuna (*Katsuwonus pelamis*, $r = 0.58$, $p < 0.1$) and South American pilchard (*Sardinops sagax*, $r = 0.69$, $p < 0.1$). These findings are consistent with previous studies showing that tuna and small pelagic species respond sensitively to productivity fluctuations in ENSO-influenced convergence zones (Lehodey et al., 2008; Wang et al., 2018; Kim et al., 2020).

In other LMEs, catch anomalies were more strongly linked to lagged chlorophyll signals. In the Southeast U.S. Continental Shelf (LME 6), northern white shrimp (*Litopenaeus setiferus*) catch anomalies were correlated with chlorophyll anomalies from the previous year ($r = 0.64$, $p < 0.05$), reflecting the influence of environmental conditions on juvenile recruitment. A similar lagged relationship was observed for Japanese jack mackerel (*Trachurus japonicus*) in the Sea of Japan (LME 50, $r = 0.57$, $p < 0.1$), potentially reflecting sensitivity to prior-year productivity during early life stages (Diop et al., 2007; Delancey et al., 1994; Takahashi et al., 2022). These results show the potential of incorporating chlorophyll-based forecasts into



fishery prediction frameworks. They also highlight the importance of accounting for species-specific life histories and ecological mechanisms when evaluating forecast performance across diverse ecosystems.

4 Conclusion and Discussion

This study demonstrates that deep learning models can achieve skillful marine ecosystem prediction in global LMEs while capturing physically interpretable signals underlying chlorophyll variability. The model successfully reproduced known ocean-climate processes, including ENSO-driven productivity fluctuations and wintertime reemergence mechanisms, suggesting that statistical learning can internalize aspects of coupled physical-biogeochemical dynamics from training data. Systematic sensitivity analyses further showed that successful data-driven ecosystem prediction requires careful consideration of both model architecture (e.g., kernel size, activation functions) and input data characteristics (e.g., horizontal resolution, log-transformation, and variable selection).

Comparisons with an ESM-based dynamical prediction system revealed regional differences in forecast skill that provide insight into the observability of marine ecosystem drivers. The deep learning model excelled in regions dominated by large-scale climate variability, where surface signals of coupled physical-biogeochemical interactions are well captured by satellite observations. However, performance limitations in eastern boundary systems highlighted the challenges of predicting large coastal ecosystems strongly influenced by subsurface processes that may not be consistently detectable at the surface. These findings emphasize that forecast skill depends not only on model design but also on the extent to which key ecological drivers are represented in available data.

The demonstrated links between predicted chlorophyll anomalies and fish catch variability provide initial evidence for the practical utility of biogeochemical forecasting in marine resource management. Statistically significant correlations were found for both contemporaneous responses (skipjack tuna, sardines) and lagged responses (shrimp, jack mackerel), patterns consistent with known life history traits and recruitment dynamics. While these relationships were identified in only a subset of LMEs, they demonstrate the feasibility of integrating environmental forecasts into fisheries applications. Such applications will require careful consideration of species-specific ecological mechanisms and regional oceanographic contexts.

Several limitations should also be acknowledged. The model treated LMEs as independent units, potentially overlooking cross-basin connectivity and anomaly propagation that could enhance predictive skill across regional boundaries. While sensitivity tests showed that surface chlorophyll anomalies effectively captured subsurface variability, other key physical drivers, such as wind stress, mixed-layer depth, and vertical nutrient gradients, were not systematically evaluated. The superior performance of dynamical models in some upwelling systems suggests that certain process-based parameterizations may be difficult to replicate through surface-based statistical learning alone. Future research should address these limitations by incorporating additional physical variables and exploring architectures that retain spatial context, such as encoder-decoder frameworks or graph-based networks, to better represent cross-basin connectivity and process-dominated systems. Moreover,



hybrid frameworks that combine machine learning with dynamical simulations, leveraging expanding Earth observation archives, offer a promising path toward transparent, flexible, and operational biogeochemical forecasting systems capable of supporting adaptive, climate-informed marine resource management.

Code and data availability

The code for the deep learning model and training procedures is available at Zendo: <https://doi.org/10.5281/zenodo.17614507> (Park et al., 2025). All observational datasets used are publicly available: satellite chlorophyll from NASA Ocean Biology Processing Group (SeaWiFS and MODIS, <https://oceandata.sci.gsfc.nasa.gov/directdataaccess/Level-3%20Mapped>), sea surface temperature from NOAA OISSTv2 (<https://www.ncei.noaa.gov/products/optimum-interpolation-sst>), and fish catch data from Sea Around Us project (<https://www.seaaroundus.org/data/#/lme>). CMIP6 simulations are accessible via Earth System Grid Federation (<https://aims2.llnl.gov/search/cmip6/>). GFDL-ECDA reanalysis data may be requested from JYP.

Author contributions

JYP and YGH conceived and designed the study. JSP and JHK developed the methodology. JSP wrote the original draft and performed the investigation with WJJ. All authors contributed to the writing.

Competing interests

The authors have no competing interests to declare.

Financial support

This work was supported by the National Research Foundation of Korea (NRF) (Grant No. RS-2025-02263830 and No. RS-2025-00564442).



References

- 395 Amorim, F. d. L. L. d., Rick, J., Lohmann, G., and Wiltshire, K. H.: Evaluation of machine learning predictions of a highly resolved time series of chlorophyll-a concentration, *Appl. Sci.*, 11, 7208, 2021.
- Balaji, V., Couvreur, F., Deshayes, J., Gautrais, J., Hourdin, F., and Rio, C.: Are general circulation models obsolete?, *Proc. Natl. Acad. Sci. USA*, 119, 10, 10.1073/pnas.2202075119, 2022.
- Balmaseda, M. A., McAdam, R., Masina, S., Mayer, M., Senan, R., de Bosissés, E., and Gualdi, S.: Skill assessment of seasonal forecasts of ocean variables, *Front. Mar. Sci.*, 11, 1380545, ARTN 1380545, 10.3389/fmars.2024.1380545, 2024.
- 400 Beal, L. M. and Bryden, H. L.: The velocity and vorticity structure of the Agulhas Current at 32 S, *J. Geophys. Res. Oceans*, 104, 5151–5176, 1999.
- Biswas, S. and Sinha, M.: Performances of deep learning models for Indian Ocean wind speed prediction, *Model. Earth Syst. Environ.*, 7, 809–831, 2021.
- 405 Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, *Science*, 359, eaam8328, 10.1126/science.aam8328, 2018.
- Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The effective number of spatial degrees of freedom of a time-varying field, *J. Clim.*, 12, 1990–2009, Doi 10.1175/1520-0442(1999)012<1990:Tenosd>2.0.Co;2, 1999.
- 410 Capotondi, A., Jacox, M., Bowler, C., Kavanaugh, M., Lehoudey, P., Barrie, D., Brodie, S., Chaffron, S., Cheng, W., Dias, D. F., Eveillard, D., Guidi, L., Iudicone, D., Lovenduski, N. S., Nye, J. A., Ortiz, I., Pirhalla, D., Buil, M. P., Saba, V., Sheridan, S., Siedlecki, S., Subramanian, A., de Vargas, C., Di Lorenzo, E., Doney, S. C., Hermann, A. J., Joyce, T., Merrifield, M., Miller, A. J., Not, F., and Pesant, S.: Observational Needs Supporting Marine Ecosystems Modeling and Forecasting: From the Global Ocean to Regional and Coastal Systems, *Front. Mar. Sci.*, 6, 623, 10.3389/fmars.2019.00623, 2019.
- 415 Cen, H. B., Jiang, J. H., Han, G. Q., Lin, X. Y., Liu, Y., Jia, X. Y., Ji, Q. Y., and Li, B.: Applying Deep Learning in the Prediction of Chlorophyll-a in the East China Sea, *Remote Sens.*, 14, 5461, 10.3390/rs14215461, 2022.
- Delancey, L. B., Jenkins, J. E., and Whitaker, J. D.: Results of long-term, seasonal sampling for *Penaeus* postlarvae at Breach Inlet, South Carolina., *Fish. Bull.*, 92, 633–640, 1994.
- 420 Diop, H., Keithly, W. R., Kazmierczak, R. F., and Shaw, R. F.: Predicting the abundance of white shrimp (*Litopenaeus setiferus*) from environmental parameters and previous life stages, *Fish. Res.*, 86, 31–41, 10.1016/j.fishres.2007.04.004, 2007.
- Esaias, W. E., Abbott, M. R., Barton, I., Brown, O. B., Campbell, J. W., Carder, K. L., Clark, D. K., Evans, R. H., Hoge, F. E., Gordon, H. R., Balch, W. M., Letelier, R., and Minnett, P. J.: An overview of MODIS capabilities for ocean science observations, *IEEE Trans. Geosci. Remote Sens.*, 36, 1250–1265, Doi 10.1109/36.701076, 1998.
- 425 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, 10.5194/gmd-9-1937-2016, 2016.
- 430 Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Höglberg, P., Linder, S., Mackenzie, F. T., Moore, B. I. I., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W.: The global carbon cycle: A test of our knowledge of earth as a system, *Science*, 290, 291–296, DOI 10.1126/science.290.5490.291, 2000.
- Falkowski, P. G., Barber, R. T., and Smetacek, V. V.: Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science*, 281, 200–207, 10.1126/science.281.5374.200, 1998.
- 435 Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L. Q.: Ocean biogeochemical modelling, *Nat. Rev. Methods Primers*, 2, 76, 10.1038/s43586-022-00154-2, 2022.
- Fiedler, P. C.: Environmental change in the eastern tropical Pacific Ocean: review of ENSO and decadal variability, *Mar.*



- Ecol. Prog. Ser., 244, 265–283, DOI 10.3354/meps244265, 2002.
- 440 Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P.: Primary production of the biosphere: integrating terrestrial and oceanic components, *Science*, 281, 237–240, 10.1126/science.281.5374.237, 1998.
- Ham, Y. G., Kim, J. H., and Luo, J. J.: Deep learning for multi-year ENSO forecasts, *Nature*, 573, 568–572, 10.1038/s41586-019-1559-7, 2019.
- Hendrycks, D. and Gimpel, K.: Gaussian error linear units (gelus), arXiv preprint arXiv:1606.08415, 2016.
- 445 Henson, S. A., Beaulieu, C., and Lampitt, R.: Observing climate change trends in ocean biogeochemistry: when and where, *Glob. Change Biol.*, 22, 1561–1571, 10.1111/gcb.13152, 2016.
- Henson, S. A., Sarmiento, J. L., Dunne, J. P., Bopp, L., Lima, I., Doney, S. C., John, J., and Beaulieu, C.: Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity, *Biogeosciences*, 7, 621–640, DOI 10.5194/bg-7-621-2010, 2010.
- 450 Hong, Z. K., Long, D., Li, X. D., Wang, Y. M., Zhang, J. M., Hamouda, M. A., and Mohamed, M. M.: A global daily gap-filled chlorophyll-a dataset in open oceans during 2001-2021 from multisource information using convolutional neural networks, *Earth Syst. Sci. Data*, 15, 5281–5300, 10.5194/essd-15-5281-2023, 2023.
- Immas, A., Do, N., and Alam, M. R.: Real-time in situ prediction of ocean currents, *Ocean Eng.*, 228, 108922, ARTN 108922, 10.1016/j.oceaneng.2021.108922, 2021.
- 455 Jeon, W., Park, J. Y., Stock, C. A., Dunne, J. P., Yang, X. S., and Rosati, A.: Mechanisms driving ESM-based marine ecosystem predictive skill on the east African coast, *Environ. Res. Lett.*, 17, 9, 10.1088/1748-9326/ac7d63, 2022.
- Keerthi, M. G., Prend, C. J., Aumont, O., and Lévy, M.: Annual variations in phytoplankton biomass driven by small-scale physical processes, *Nat. Geosci.*, 15, 1027–1033 10.1038/s41561-022-01057-3, 2022.
- Kim, J., Na, H., Park, Y. G., and Kim, Y. H.: Potential predictability of skipjack tuna (*Katsuwonus pelamis*) catches in the Western Central Pacific, *Sci. Rep.*, 10, 3193, 10.1038/s41598-020-59947-8, 2020.
- 460 Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H. M., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*, 17, 3439–3470, 10.5194/bg-17-3439-2020, 2020.
- 465 Lehodey, P., Senina, I., and Murtugudde, R.: A spatial ecosystem and populations dynamics model (SEAPODYM) - Modeling of tuna and tuna-like populations, *Prog. Oceanogr.*, 78, 304–318, 10.1016/j.pcean.2008.06.004, 2008.
- Liu, T. C., Yu, G. C., Kwok, H. Y., Xue, R. Z., He, D., and Liang, W. Z.: Enhancing tree-based machine learning for chlorophyll-a prediction in coastal seawater through spatiotemporal feature integration, *Mar. Environ. Res.*, 209, 107170, 10.1016/j.marenvres.2025.107170, 2025.
- 470 Lundberg, S. M., Erion, G. G., and Lee, S.-I.: Consistent individualized feature attribution for tree ensembles, arXiv preprint arXiv:1802.03888, 2018.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S. I.: From Local Explanations to Global Understanding with Explainable AI for Trees, *Nat Mach Intell*, 2, 56–67, 10.1038/s42256-019-0138-9, 2020.
- 475 McClain, C. R., Cleave, M. L., Feldman, G. C., Gregg, W. W., Hooker, S. B., and Kuring, N.: Science quality SeaWiFS data for global biosphere research, *Sea Technol.*, 39, 10–16, 1998.
- 480 Mignot, A., Claustre, H., Cossarini, G., d’Ortenzio, F., Gutknecht, E., Lamouroux, J., Lazzari, P., Perruche, C., Salon, S., and Sauzède, R.: Using machine learning and Biogeochemical-Argo (BGC-Argo) floats to assess biogeochemical models and optimize observing system design, *Biogeosciences*, 20, 1405–1422, 2023.
- Park, J. S., Park, J. Y., Ham, Y. G., Kim, J. H., and Jeon, W.: Deep learning-based chlorophyll prediction: comparison with a dynamic model and applications to fish catch forecasting. Zenodo. <https://doi.org/10.5281/zenodo.17614507>, 2025.



- 485 Park, J. Y., Dunne, J. P., and Stock, C. A.: Ocean Chlorophyll as a Precursor of ENSO: An Earth System Modeling Study, *Geophys. Res. Lett.*, 45, 1939–1947, 10.1002/2017gl076077, 2018a.
- 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, 10.1126/science.aav6634, 2019.
- 490 Park, J. Y., Stock, C. A., Yang, X. S., Dunne, J. P., Rosati, A., John, J., and Zhang, S. Q.: Modeling Global Ocean Biogeochemistry With Physical Data Assimilation: A Pragmatic Solution to the Equatorial Instability, *J. Adv. Model. Earth Syst.*, 10, 891–906, 10.1002/2017ms001223, 2018b.
- Pauly, D. and Zeller, D.: Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining, *Nat. Commun.*, 7, 10244, 10.1038/ncomms10244, 2016.
- Pennington, J. T., Mahoney, K. L., Kuwahara, V. S., Kolber, D. D., Calienes, R., and Chavez, F. P.: Primary production in the eastern tropical Pacific: A review, *Prog. Oceanogr.*, 69, 285–317, 10.1016/j.pocean.2006.03.012, 2006.
- 495 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, *Nature*, 566, 195–204, 10.1038/s41586-019-0912-1, 2019.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily high-resolution-blended analyses for sea surface temperature, *J. Clim.*, 20, 5473–5496, 10.1175/2007jcli1824.1, 2007.
- 500 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, *Front. Mar. Sci.*, 10, 1077623, 10.3389/fmars.2023.1077623, 2023.
- Sasai, Y., Richards, K. J., Ishida, A., and Sasaki, H.: Spatial and temporal variabilities of the chlorophyll distribution in the northeastern tropical Pacific: The impact of physical processes on seasonal and interannual time scales, *J. Mar. Syst.*, 96–97, 24–31, 10.1016/j.jmarsys.2012.01.014, 2012.
- 505 Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., and Dunne, J.: Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6, *Curr. Clim. Change Rep.*, 6, 95–119, 2020.
- Siegel, D. A., DeVries, T., Cetinic, I., and Bisson, K. M.: Quantifying the Ocean's Biological Pump and Its Carbon Cycle Impacts on Global Scales, *Annu. Rev. Mar. Sci.*, 15, 329–356, 10.1146/annurev-marine-040722-115226, 2023.
- 510 Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., and Yang, X.: Robust skill of decadal climate predictions (vol 3, pg 15, 2020), *npj Clim. Atmos. Sci.*, 3, 1, 2020.
- 515 Song, T., Jiang, J. Y., Li, W., and Xu, D. Y.: A Deep Learning Method With Merged LSTM Neural Networks for SSHA Prediction, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 13, 2853–2860, 10.1109/Jstars.2020.2998461, 2020.
- Stock, C. A., Dunne, J. P., and John, J. G.: Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical-biological model, *Prog. Oceanogr.*, 120, 1–28, 10.1016/j.pocean.2013.07.001, 2014.
- 520 Sweeney, C., Gnanadesikan, A., Griffies, S. M., Harrison, M. J., Rosati, A. J., and Samuels, B. L.: Impacts of shortwave penetration depth on large-scale ocean circulation and heat transport, *J. Phys. Oceanogr.*, 35, 1103–1119, Doi 10.1175/Jpo2740.1, 2005.
- Takahashi, M., Sassa, C., Kitajima, S., Yoda, M., and Tsukamoto, Y.: Linking environmental drivers, juvenile growth, and recruitment for Japanese jack mackerel *Trachurus japonicus* in the Sea of Japan, *Fish. Oceanogr.*, 31, 70–83, 2022.
- 525 Volk, T. and Hoffert, M. I.: Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes, in: *The carbon cycle and atmospheric CO₂: Natural variations Archean to present* edited by: Sundquist, E. T., and Broecker, W. S., *Geophysical Monograph Series*, 32, American Geophysical Union, Washington, DC, 99–110, 10.1029/GM032p0099, 1985.
- Wang, J. T., Chen, X. J., Staples, K. W., and Chen, Y.: The skipjack tuna fishery in the west-central Pacific Ocean: applying neural networks to detect habitat preferences, *Fish. Sci.*, 84, 309–321, 10.1007/s12562-017-1161-6, 2018.



- 530 Xiao, C. J., Chen, N. C., Hu, C. L., Wang, K., Xu, Z. W., Cai, Y. P., Xu, L., Chen, Z. Q., and Gong, J. Y.: A spatiotemporal
 deep learning model for sea surface temperature field prediction using time-series satellite data, *Environ. Model.*
Softw., 120, 104502, 10.1016/j.envsoft.2019.104502, 2019.
- Yao, L. L., Wang, X. P., Zhang, J. H., Yu, X., Zhang, S. C., and Li, Q.: Prediction of Sea Surface Chlorophyll-a
 Concentrations Based on Deep Learning and Time-Series Remote Sensing Data, *Remote Sens.*, 15, 4486,
 535 10.3390/rs15184486, 2023.
- Yu, B. W., Xu, L. L., Peng, J. H., Hu, Z. Z., and Wong, A.: Global chlorophyll-a concentration estimation from moderate
 resolution imaging spectroradiometer using convolutional neural networks, *J. Appl. Remote Sens.*, 14, 17,
 10.1117/1.Jrs.14.034520, 2020.
- 540 Zhang, S., Harrison, M. J., Rosati, A., and Wittenberg, A.: System design and evaluation of coupled ensemble data
 assimilation for global oceanic climate studies, *Mon. Weather Rev.*, 135, 3541–3564, 10.1175/Mwr3466.1, 2007.