*Corresponding Author's Email: yashnilmohanty1@gmail.com
Proceedings of the World Conference on Climate Change and Global Warming
Vol. 2, Issue. 2, 2025, pp. 14-38
DOI: https://doi.org/10.33422/ccgconf.v2i2.1619

ISSN: 3030-0703 online

Physics-Guided Neural Forecasts of Nearshore Harmful Algal Blooms in the California Current System

Yashnil Mohanty
Westmont High School, Campbell, CA, United States

Abstract

Harmful algal blooms (HABs) are increasing in frequency, duration, and extent along the California coast, driven by climate variability, nutrient enrichment, and complex physical–biogeochemical interactions. Forecasting HAB development and spread remains a challenge, especially in Eastern Boundary Upwelling Systems where advection, stratification, and episodic river inputs strongly shape bloom dynamics. Existing approaches often trade physical realism for statistical flexibility, limiting generalization across bloom regimes. We present a physics-informed deep learning framework for nearshore chlorophyll-a forecasting in the California Current System, integrating multi-sensor satellite products, atmospheric and ocean reanalysis fields, and static geospatial predictors. Three architectures are evaluated: a convolutional long short-term memory network (ConvLSTM), a Temporal Fusion Transformer (TFT), and a physics-informed ConvLSTM (PINN) incorporating the two-dimensional advection–diffusion equation as a soft training constraint. A multi-year, 4 km-resolution dataset (2003–2021) is processed via a tailored feature engineering pipeline with quality-controlled gap-filling, rolling statistics, lagged predictors, and climatology-based anomalies. Models are assessed with strict spatiotemporal cross-validation, emphasizing spatial fidelity, bloom footprint representation, and predictor interpretability. Post-hoc explainability analyses identify key environmental drivers consistent with known upwelling–bloom linkages in the region. We present comparative skill assessments, spatial bias analyses, and predictor attribution results, highlighting the advantages and trade-offs of adding physical constraints to coastal HAB forecasting models. This work delivers a scalable and transferable methodology with direct implications for ecosystem management, fisheries, and public health.

Keywords: harmful algal blooms, spatiotemporal modeling, physics-informed neural networks, ocean forecasting, remote sensing

1. Introduction

Harmful algal blooms (HABs) are ecological disturbances caused by the rapid accumulation of algae in aquatic systems, often triggered by excess nutrients, warming surface waters, and stratification (Litchman, 2022; Ralston and Moore, 2020). These events are increasingly common across global coastal zones, with rising frequency and severity driven by anthropogenic nutrient enrichment and climate change. Along the California coast, blooms formed by taxa such as *Karenia brevis*, *Pseudo-nitzschia*, and cyanobacteria pose serious risks to ecosystems and human health (Ryan et al., 2017; Thorel et al., 2017). Impacts range from marine wildlife mortality and shellfish toxicity to respiratory irritation, drinking water contamination, and economic disruption to coastal fisheries and tourism (Grattan et al. 2016).

The societal and ecological stakes of HABs have motivated efforts to develop early warning systems that can forecast bloom development and spatial spread (Moore et al., 2008). Timely predictions allow coastal managers, shellfish farmers, and water utilities to take proactive measures—issuing harvest closures, deploying mitigation technologies, or adjusting water treatment processes—to reduce harm and economic loss (Molares-Ulloa et al., 2024). However, despite the pressing need for reliable forecasting tools, existing HAB prediction systems are limited in spatial coverage, generalizability, and mechanistic fidelity (Marrone et al., 2023). In this study, we focus on chlorophyll-*a* as a proxy for bloom intensity due to its availability from satellite remote sensing and its widespread use in HAB surveillance (Demiray et al., 2025).

Traditional HAB forecasting frameworks fall into two major categories: process-based simulation and statistical modeling. Process-based models are grounded in ecological theory and use mechanistic representations of phytoplankton dynamics, including nutrient uptake, growth rates, light limitation, and advection by ocean currents. While these models can reproduce known bloom mechanisms in data-rich regions, they are sensitive to parameter uncertainty, often require species-specific tuning, and are computationally expensive (Franks, 2025). They also struggle to generalize across different bloom types or regions with limited observational infrastructure (McGillicuddy, 2010).

Statistical and machine learning (ML) models, on the other hand, can identify empirical correlations between environmental drivers and bloom metrics using historical data (Demiray et al., 2025). These approaches range from generalized linear models and random forests to recurrent neural networks. While easier to train and often more scalable, statistical models may overfit to local patterns and break down under non-stationary or sparse conditions. They also typically lack mechanistic interpretability and may produce unrealistic spatial outputs that violate known physical constraints (Anderson et al., 2019; Park et al., 2024).

Recent advances in deep learning and hybrid modeling have opened new frontiers in ecological forecasting, particularly in data-sparse or dynamically complex systems. Physics-informed neural networks (PINNs) and other constraint-aware architectures have been used to incorporate governing equations—such as the advection-diffusion equation or shallow-water dynamics—directly into the training objective of neural models (Raissi et al., 2019). These hybrid approaches have shown promise in applications ranging from pollutant plume tracking and ocean circulation to solar irradiance and hydrologic forecasting, enabling models to learn from limited data while remaining grounded in physical plausibility.

Within the growing body of HAB forecasting literature, machine learning techniques have been increasingly applied to predict bloom dynamics—most commonly through time series forecasting or binary classification based on satellite-derived chlorophyll data. Hill et al. developed *HABNet*, a hybrid system combining convolutional and recurrent neural networks

with spatiotemporal datacube representations of HAB events (Hill et al., 2020). Applied to *Karenia brevis* blooms along the Florida coast, HABNet achieved high classification accuracy and demonstrated 8-day predictive skill, though it was restricted to a single species and region. Molares Ulloa et al. introduced hybrid models for estimating shellfish harvest closures based on biotoxin thresholds, comparing advanced techniques such as BAGNET and SVM-KNN across multiple estuarine systems (Molares-Ulloa et al., 2024). Their approach improved robustness in operational classification tasks but remained focused on regulatory targets rather than mechanistic bloom drivers or spatial transport. More broadly, Cruz et al. reviewed recent advances in machine learning for HAB and shellfish contamination prediction, highlighting the growing complexity of forecasting models and increased use of multi-source inputs (Cruz et al., 2021). However, they noted persistent limitations, including the scarcity of models that account for physical ocean dynamics, generalize across space, or incorporate process-based knowledge. Addressing this gap, Gonzalez proposed a trait-based ecological framework for predicting HABs, arguing for the integration of species-specific traits into statistical and mechanistic models to improve forecasts under changing environmental conditions (Litchman, 2022). Despite this momentum, few existing approaches explicitly incorporate fluid transport physics or evaluate predictive skill on spatial spread—two elements that are critical for delivering actionable HAB forecasts to coastal managers.

This study builds on the growing intersection of machine learning and geophysical modeling by testing whether physically constrained neural networks can improve chlorophyll-*a* forecasting skill in coastal California. We focus on a multi-year, multi-sensor dataset clipped to a ~10-mile nearshore strip and evaluate model skill on both pixelwise accuracy and bloom-scale spatial structure across multiple chlorophyll regimes. Our central hypothesis is that embedding coastal advection–diffusion physics into a deep spatiotemporal network—via residual penalties applied to log-chlorophyll predictions—will improve 8-day forecast skill, particularly for high-chlorophyll events and under combined temporal–spatial distribution shifts. We expect that the physics-informed approach will (i) reduce false positives and unrealistic alongshore spread, (ii) better reproduce observed bloom orientation, scale, and spectral energy distribution, and (iii) maintain competitive performance in low-chlorophyll conditions. By coupling these improvements with predictor attribution analyses, we also aim to identify physically meaningful drivers (e.g., wind stress curl, Kd490, SST, river proximity) and quantify how their influence varies seasonally and regionally within the California Current System. In doing so, we seek to deliver a scalable, interpretable, and transferable forecasting framework that advances the state of operational coastal bloom prediction.

2. Methods

2.1 Study Domain

Our study focuses on a ~10-mile (~16 km) nearshore strip along the California coast, spanning from the Southern California Bight to the Oregon spillover zone (Figure 1). This coastal band encompasses diverse marine ecosystems influenced by wind-driven upwelling, riverine nutrient discharge, mesoscale eddies, and frontal dynamics (Anderson et al., 2008). The domain is defined using a binary coastal stripe mask applied to Level-3 mapped satellite composites, excluding land and deep offshore waters to isolate HAB-relevant areas. The mask includes only ocean pixels whose geodesic distance to the shoreline is ≤16 km, computed from Natural Earth coastlines in UTM projection (EPSG:32610) and eroded slightly to prevent land bleed.

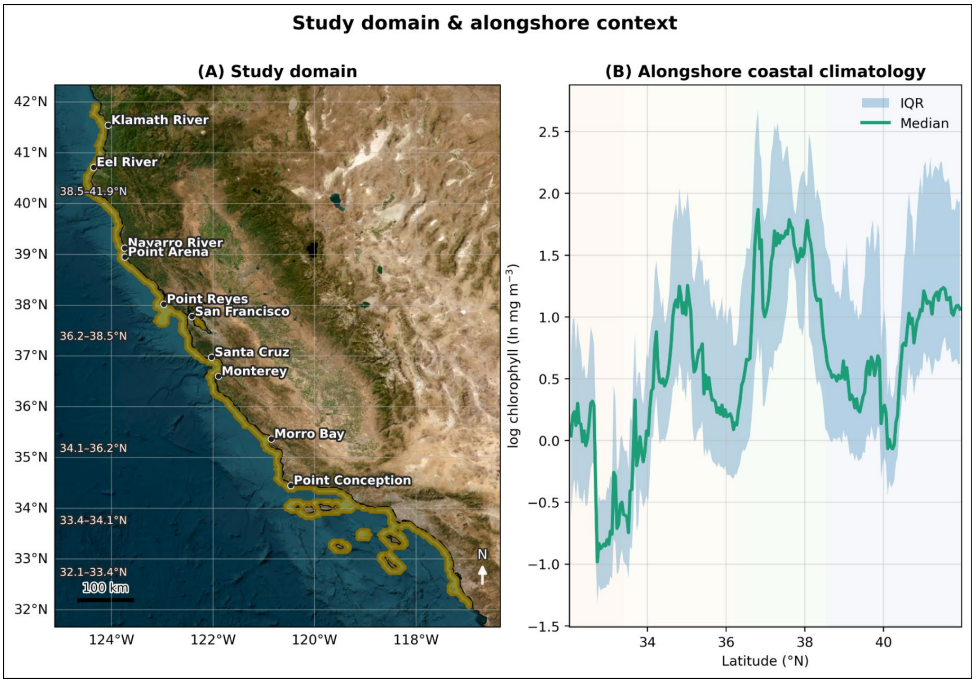

Figure 1. Study domain and coastal mask: ~10-mile nearshore strip, grid resolution (~4km), and key regions (SoCal Bight, Monterey Bay, Northern CA).

The resulting analysis grid contains 240 x 240 cells at ~4 km nominal resolution, with analysis restricted to valid ocean pixels. The dataset spans 18.6 years (January 2003–mid-2021), comprising 851 non-overlapping 8-day time steps (Δt = 691,200s). This period captures seasonal, interannual, and decadal variability, including El~Niño–Southern Oscillation (ENSO) events, marine heatwaves, and shifts in upwelling intensity (Jacox et al. 2016). At the native resolution, the grid resolves large-scale bloom patterns, shelf gradients, and major upwelling plumes, though it does not fully capture submesoscale filaments (<2 km) (Le Traon, 2015). Sub-grid variability is partially represented through static and dynamic predictors such as distance-to-river, wind stress curl, and sea surface height gradients.

The spatiotemporal coverage is sufficient for operational coastal forecasting and for testing model generalization across a range of physical regimes. Key ecological and management zones within the mask include the Monterey Bay National Marine Sanctuary, nutrient-rich estuaries such as the Eel and Navarro rivers, and the Southern California Bight. This consistent, multi-decadal observational record, paired with ancillary reanalysis fields, provides the foundation for physics-informed machine learning experiments aimed at predicting and understanding harmful algal bloom dynamics along the California coast.

2.2 Data Sources and Transformation

Table 1 lists the core predictors with their units, data sources, and physical relevance. Selection was guided by physical and biogeochemical processes known to influence phytoplankton dynamics, including nutrient delivery, light availability, wind-driven mixing, and coastal circulation.

Ocean-colour observations are obtained from NASA Earthdata's MODIS-Aqua Level-3 mapped products at ~4 km resolution ("4.6 km at equator"), provided as non-overlapping 8-day composites to reduce cloud-related gaps while retaining weekly-scale bloom dynamics. Variables include: chlorophyll-*a* concentration (natural log transform with a positive floor ε=0.056616 mg m^-3; a binary floor flag channel is retained for censoring awareness), daytime sea surface temperature (from the 11 μm thermal IR band), normalized fluorescence

line height (nFLH), and the diffuse attenuation coefficient at 490 nm (Kd490). The log transform stabilizes variance, deemphasizes heavy-tailed extremes, and improves statistical robustness.

Table 1. Predictor variables, data sources, and relevance for HAB forecasting

Predictor (units)	**Data source**	**Relevance**
Chlorophyll-a concentration (mg m^{-3})	**NASA Earthdata** (MODIS-Aqua L3)	Primary biomass proxy; basis for bloom detection and derived chl features.
Fluorescence Line Height, nFLH (mW cm^{-2} μm^{-1} sr^{-1})	**NASA Earthdata** (MODIS-Aqua L3)	Proxy for chlorophyll fluorescence/photosynthetic activity; helps flag high-biomass events.
Diffuse attenuation Kd_{490} (m^{-1})	**NASA Earthdata** (MODIS-Aqua L3)	Water clarity/light penetration; constrains euphotic depth and growth potential.
10 m wind (u) (m s^{-1})	**ERA5**	Zonal wind; along-shore component drives upwelling/downwelling.
10 m wind (v) (m s^{-1})	**ERA5**	Meridional wind; cross-shore component affects onshore/offshore transport.
Wind speed (m s^{-1})	**ERA5 (derived from u10,v10)**	Turbulent mixing/stratification; modulates light and nutrient supply.
Wind stress magnitude $\|\tau\|$ (N m^{-2})	**ERA5 (derived)**	Ekman transport/upwelling forcing; useful for identifying upwelling-favorable periods.
Surface downward shortwave radiation (W m^{-2})	**ERA5**	Light availability for photosynthesis; interacts with stratification.
Total precipitation (mm day^{-1})	**ERA5**	Freshwater/stratification pulses; potential nutrient and turbidity impacts nearshore.
2 m air temperature (K)	**ERA5**	Atmosphere–ocean heat flux proxy; influences SST and stratification.
2 m dewpoint temperature (K)	**ERA5**	Humidity/latent heat flux; coupled to mixing and cloud cover.
Eastward surface current, uo (m s^{-1})	**CMEMS (GLORYS)**	Advection of biomass and nutrients; along-shore transport.
Northward surface current, vo (m s^{-1})	**CMEMS (GLORYS)**	Advection; cross-frontal transport and retention.
Surface current speed (m s^{-1})	**CMEMS (derived)**	Net transport intensity; dispersion/retention of blooms.
Horizontal divergence (s^{-1})	**CMEMS (derived)**	Convergence/divergence zones; nutrient supply vs. retention.

Predictor (units)	Data source	Relevance
Relative vorticity (s^{-1})	**CMEMS (derived)**	Eddies/fronts; submesoscale trapping and bloom confinement.
Sea surface height, SSH (m)	**CMEMS (GLORYS)**	Large-scale circulation; mesoscale features and steric height.
SSH gradient magnitude |∇SSH| (m m^{-1})	**CMEMS (derived)**	Front intensity; proxies for geostrophic shear and frontal zones.
Sea surface salinity, SSS (PSU)	**CMEMS (GLORYS)**	Freshwater inputs/stratification; river plumes and buoyancy effects.
Sea surface temperature, SST (°C)	**CMEMS (GLORYS; θ)**	Thermal control on growth rates and stratification; fronts.

Atmospheric drivers are sourced from ECMWF's ERA5 reanalysis via the Copernicus Climate Data Store, aggregated to 8-day means aligned with MODIS composite windows. Variables include 10 m zonal/meridional winds, 2 m air and dewpoint temperatures, total precipitation, and mean surface downward shortwave radiation. Wind speed is computed from vector components, and wind stress is estimated using the quadratic drag law:

$$\boldsymbol{\tau} = \rho_a C_D \parallel \boldsymbol{u}_{10} \parallel \cdot \boldsymbol{u}_{10}$$

where $\boldsymbol{u}_{10}$is the 10 m wind vector, ρ_a is air density, and C_D is the bulk drag coefficient. Wind stress curl, a proxy for upwelling, is computed in projected metric coordinates using centered finite differences with $cos(\varphi)$ scaling to preserve unit consistency (Large and Pond, 1981).

Ocean state variables come from the CMEMS GLORYS Global Ocean Physics Reanalysis, providing daily sea surface height (η), potential temperature (θ_o), salinity (S_o), and surface velocity components (u_o, v_o) at ~8 km resolution. Fields are bilinearly interpolated to the 4 km MODIS grid, lightly smoothed to reduce aliasing, and then used to derive current speed, horizontal divergence, relative vorticity, and sea surface height gradient magnitude:

$$|\nabla\eta| = \sqrt{\left(\frac{\partial\eta}{\partial x}\right)^2 + \left(\frac{\partial\eta}{\partial y}\right)^2}$$

with derivatives computed using centered finite differences and metric-aware grid spacing. These features capture mesoscale fronts and geostrophic transport structures relevant to bloom advection and retention (Morrow et al., 2004).

From both observed and reanalysis variables, we compute rolling means and standard deviations over 24-day and 40-day windows, monthly anomalies relative to a long-term training-set climatology (to prevent leakage), and lagged versions (1–3 timesteps) of chlorophyll-*a* and key covariates. Riverine influence is quantified via the geodesic distance to the nearest river mouth and an exponential decay index exp(-*d*/*l*), where *d* is offshore distance and *l* is a prescribed nutrient decay length scale. Static spatial features include latitude, distance-to-coast, and river-rank indices.

All predictors are quality-controlled, collocated to the coastal mask, standardized using training-period statistics, and stored as aligned spatiotemporal arrays for model ingestion.

2.3 Feature Engineering

All predictors are prepared as aligned spatiotemporal arrays on the masked ~4 km coastal grid, quality-controlled, and standardized using training-period statistics to prevent information leakage. A binary *valid_mask* channel is preserved throughout to ensure only valid ocean pixels contribute to training loss or metrics.

Ocean-colour variables (chlorophyll-*a*, Kd490, nFLH) are gap-filled within the coastal stripe to address missing observations from cloud contamination in MODIS imagery. The imputation method combines conservative temporal and spatial interpolation with physical context from wind direction, leveraging the tendency for bloom advection and dispersion to align with prevailing wind patterns. During nearest-neighbour filling of missing pixel medians, donor candidates are preferentially searched along the temporally averaged wind vector for that time block, with along-wind distances weighted more heavily than cross-wind distances. This directional bias preserves the anisotropic transport structure characteristic of upwelling-driven coastal systems. Donors are restricted to contiguous in-mask ocean pixels within a 3-pixel radius; imputation is skipped if the local observed fraction is < 40%.

For chlorophyll-*a* and Kd490, imputation is performed in log space, merging multi-year pixel medians with regional 8-day seasonal anomalies and applying soft caps on multiplicative departures from the baseline. nFLH, which may take negative values, is processed in linear space using pixel medians and local interquartile ranges with analogous seasonal reconstruction logic. Observed values are preserved in all cases, and observation flags are retained for traceability.

A climatology + clipping routine is implemented for robustness checks: short temporal gaps (≤ 24d) are bridged linearly, remaining holes are filled with a smoothed 8-day block climatology, and synthetic values are clipped to the observed 1st–99th percentile range.

The chlorophyll-*a* target is expressed as natural log concentration with a fixed positive floor ε=0.056616 mg m^-3; a binary "floor flag" channel is retained so the model can learn background clarity rather than treating floor-imposed values as missing. From driver fields, we derive rolling means and standard deviations over 24-day and 40-day windows, monthly anomalies relative to a long-term training-set climatology, and lagged versions (1–3 time steps) of chlorophyll-*a* + key covariates. Static spatial features include latitude, distance-to-coast, geodesic distance to the nearest river mouth, an exponential decay index $\exp(-d/l)$ parameterizing offshore nutrient influence, and river-rank indices. All spatial derivatives (e.g., wind stress curl, sea surface height gradients, current divergence/vorticity) are computed in projected metric coordinates using centered finite differences with $\cos(\varphi)$ scaling to ensure unit consistency.

To better represent the full dynamic range of coastal chlorophyll, training sequences are stratified by *linear*-space chlorophyll regime. Four strata (Background, Elevated, Bloom, Extreme) are defined by month-balanced percentiles from the training distribution, ensuring that Bloom and Extreme events are not underrepresented. Mini-batches draw approximately equal samples from each stratum, and the supervised loss is weighted inversely by bin frequency to counter the natural skew toward low-chlorophyll states.

To assess robustness to data gaps, a sensitivity analysis tested seven imputation methods (original/no imputation, zero-fill, mean-fill, median-fill, forward-fill, climatology, spatial interpolation) on validation-set ocean-color variables. The pre-trained PINN model was

evaluated on each imputed dataset without retraining to isolate imputation effects from model adaptation.

2.4 Cross-Validation

We implement a blocked, stratified spatiotemporal cross-validation scheme to assess model generalization under temporal and spatial distribution shifts, while ensuring balanced representation of chlorophyll regimes. The primary temporal split uses 2003–2015 for training, 2016–2018 for validation, and 2019–mid-2021 for testing, thereby holding out major interannual events (e.g., the 2014–2016 warm anomaly, 2019 marine heatwave) for evaluation. An alternate *k*-fold scheme with sliding 5-year windows advanced by 2 years is available for robustness testing.

For spatial evaluation, the masked coastal grid is divided into N=5 contiguous alongshore bands of approximately equal valid-pixel count, determined by projecting cell centroids into along-shelf distance coordinates and clustering to preserve latitudinal coherence. Spatial out-of-sample (OOS) experiments withhold one band for validation or testing while training on the remainder. In the *STRICT_OOS* mode, both temporal and spatial holdouts are applied simultaneously (e.g., the most recent years in the northernmost band), representing the most challenging extrapolation scenario.

We report results for (A) fixed spatial domain, temporal holdout only; (B) all years from a withheld spatial band; and (C) combined temporal and spatial holdouts. Metrics are computed in both log- and linear-space chlorophyll for all modes. Sampling within each fold is stratum-balanced by month-adjusted chlorophyll regime (Background, Elevated, Bloom, Extreme) to avoid overrepresentation of low-chlorophyll states in training or evaluation.

For operational-style forecasts, the model is retrained on the combined training + validation period (2003–2018) and evaluated on the 2019–mid-2021 holdout, reflecting the real-world constraint of predicting future blooms along a known coastline.

2.5 Model Architecture

We compare three modeling approaches for forecasting nearshore chlorophyll-*a*: a baseline convolutional long short-term memory network (ConvLSTM), a Temporal Fusion Transformer (TFT), and a physics-informed ConvLSTM (PINN). All models take as input L historical 8-day frames of multivariate predictors and output the chlorophyll-*a* field at $t + \Delta t$, with $\Delta t = 8$ days for the primary experiments (Reichstein et al., 2019).

The ConvLSTM baseline integrates convolutional layers into an LSTM framework, enabling simultaneous modeling of spatial coherence and temporal dependencies. Input tensors of shape [L, C, H, W] are processed to produce predictions in log-transformed chlorophyll space. This architecture is well-suited to tracking advective features, but is prone to smoothing sharp gradients and underestimating extreme bloom events (Amato et al., 2020).

The TFT is an attention-based, multi-horizon forecasting model adapted here for coastal geophysical grids. The masked coastal grid is flattened into a sequence of valid spatial points, with static spatial metadata (e.g., latitude, distance-to-coast, distance-to-river) appended as static features. Variable selection networks, gating layers, and temporal attention allow the TFT to handle heterogeneous covariates and capture long-range dependencies. The model is trained with a quantile regression loss at $\tau \in \{0.1, 0.5, 0.9\}$ to produce probabilistic forecasts; here we report $\tau = 0.5$ for point predictions, while probabilistic evaluation (e.g., CRPS, reliability diagrams, sharpness) uses the full quantile set (Lim et al., 2021; Pathak et al., 2022).

The PINN model augments the ConvLSTM with a physics-based loss term enforcing the 2-D advection–diffusion equation in standardized (z-score) log-chlorophyll space and including a learnable source term S(*x,y,t*) to capture unresolved biological growth and mixing (Raissi et al., 2019; Gruber et al., 2011; Marchesiello et al., 2001):

$$\frac{C^{t+\Delta t} - C^t}{\Delta t} + u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} - \kappa\nabla^2 C - S = 0$$

where C is log-transformed chlorophyll-*a*, $\mathbf{u} = (u, v)$ is the surface velocity vector from reanalysis (optionally including Ekman transport), and κ is a learnable horizontal diffusivity parameterized via the softplus function to enforce non-negativity, initialized to 25 m^2 s^-1 (Marchesiello et al., 2003). Spatial derivatives are implemented as convolutional stencils with metric scaling, applied only to valid ocean pixels. Zero-gradient (Neumann) boundaries are applied along coast-adjacent mask edges, and open-radiation boundaries are applied offshore (Marchesiello et al., 2001; Marchesiello et al., 2003).

Training proceeds in two phases: (1) a supervised phase using a class-weighted Huber loss on log C with stratified mini-batches to ensure balanced representation of Background, Elevated, Bloom, and Extreme regimes; and (2) a physics-augmented phase in which the physics residual loss is introduced with a ramped weight λ increasing from 0.0 to 1.0 over 5–10 epochs after the validation RMSE plateaus (Raissi et al., 2019). The combined loss encourages forecasts that are both statistically accurate and dynamically consistent with coastal advection–diffusion, reducing displacement errors, constraining unrealistic bloom spread, and improving spatial orientation fidelity.

Uncertainty quantification methods were implemented for the PINN (Monte Carlo Dropout, ensemble training, quantile regression) to complement the TFT's probabilistic outputs, though comprehensive evaluation remains a direction for future work. The primary experiments reported here focus on deterministic point predictions.

Model training was performed on consumer-grade hardware (Apple Silicon or equivalent), with full training typically requiring 3–6 hours for the PINN over 30–40 epochs. Inference on a single 8-day forecast takes <10 seconds on CPU and substantially faster on GPU acceleration, enabling rapid operational deployment. The physics residual computation adds minimal overhead (~10–15% inference time), making real-time forecasting feasible for the coastal domain.

2.6 Hyperparameter Selection and Ablation

To determine optimal hyperparameters for the physics-informed model, we conducted systematic ablation studies on the physics loss weight (λ) and initial diffusivity (κ). The physics loss weight λ controls the relative importance of the advection-diffusion constraint versus the supervised prediction loss, while κ sets the initial horizontal diffusivity that governs bloom spread and smoothing.

For the lambda ablation, we tested seven values ($\lambda \in \{0.0, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0\}$) with κ fixed at 25 m^2 s^-1. The $\lambda = 0.0$ case corresponds to a pure ConvLSTM baseline without physics constraints. Each configuration was trained for 30 epochs using the trained set, with validation performance monitored on the validation period to prevent information leakage from the test set.

For the kappa ablation, we fixed λ at the value selected from the lambda ablation and tested $\kappa \in \{1, 5, 10, 25, 50, 100, 500\}$ m^2 s^-1. These values span typical mesoscale diffusivity

ranges reported for eastern boundary upwelling systems. The same training and validation procedure was applied as in the lambda ablation.

In both studies, models were trained with the same architecture, data splits, and training procedures (stratified sampling, class-weighted Huber loss) to ensure fair comparison. The validation RMSE (log space) was used as the primary selection criterion, with additional consideration given to spatial fidelity metrics and physics residual magnitudes.

2.7 Evaluation Metrics

Model performance is quantified using both continuous (regression) and event-based (categorical) metrics, evaluated on the masked coastal grid. Regression skill is assessed in both log- and linear-space chlorophyll-*a* using root-mean-square error (RMSE), mean absolute error (MAE) (Willmott and Matsuura, 2005), Kling–Gupta efficiency (KGE) (Gupta et al., 2009), and rank- and linear-correlation coefficients (Spearman's ρ, Pearson's *r*). These metrics capture bias, variance, and correlation structure relative to observations. Frames with < X% valid coverage (post-mask) are down-weighted in domain-mean metrics, and a time–latitude *data coverage Hovmöller* is reported for context.

Categorical bloom-detection skill is computed by thresholding the *linear* chlorophyll field at a default bloom criterion of 5 mg m^-3, with sensitivity tests at 3 and 10 mg m^-3. Metrics include hit rate (probability of detection), false alarm ratio, precision, recall, F1 score, receiver operating characteristic area under the curve (ROC-AUC), intersection-over-union (IoU) for binary bloom masks, and centroid displacement (km) between observed and forecast bloom patches (Sokolova and Lapalme, 2009).

Spatial spread fidelity is evaluated using multiple complementary measures: normalized RMSE across the bloom mask, morphological comparison of binary bloom masks, relative scale error (bias in integral length scale), and spectral energy ratio E_pred(k) / E_obs(k) over wavenumber bands (1.0 = perfect). These metrics emphasize the model's ability to reproduce the spatial footprint, scale distribution, and dispersion characteristics of observed blooms, beyond pointwise accuracy (Arbic et al., 2012).

For probabilistic models (e.g., TFT), calibration and sharpness are assessed using the continuous ranked probability score (CRPS), reliability diagrams, and forecast interval width (IQR) [33]. For the physics-informed model, additional diagnostics include the distribution of the physics residual, the ratio of gradient norms between physics and supervised components, and the fractional contribution of the physics term to the total loss. Together, these metrics quantify not only forecast accuracy but also spread realism, uncertainty quality, and physical consistency (Raissi et al., 2019).

2.8 Predictor Importance Analysis

To interpret model predictions and quantify the influence of individual predictors, we apply post hoc explainability analyses tailored to both tabular baselines and deep spatiotemporal networks. For pixelwise gradient-boosted tree (GBDT) baselines and temporally/spatially aggregated ConvLSTM or TFT inputs, we compute Shapley additive explanations (SHAP) to attribute each predictor's contribution to forecasted chlorophyll-*a* at each timestep. Global importance is obtained by averaging absolute SHAP values across all valid coastal pixels and test-period samples (Lundberg and Lee, 2017).

For deep learning architectures (ConvLSTM, PINN), we complement SHAP—computed via a pixelwise GBDT surrogate—with direct *occlusion* and *permutation* importance on the raw

driver channels. These perturbation-based methods verify that surrogate-based attributions are consistent with the actual sensitivity of the trained network.

For physically oriented interpretation, we examine partial dependence plots (PDPs) and bivariate scatterplots for the top-ranked predictors (Friedman, 2001). PDPs quantify the marginal effect of a single predictor (e.g., wind stress curl, Kd490, SST) on predicted chlorophyll while holding other inputs at representative values, enabling identification of nonlinear thresholds, saturation behavior, and sign reversals. Scatterplots of predictor–response pairs in both raw and anomaly space are used to validate SHAP- or permutation-derived relationships against empirical data patterns, ensuring consistency between statistical attribution and observed process dynamics.

Together, these analyses yield a multi-perspective view of predictor influence—pairing model-agnostic SHAP for comparability across architectures with perturbation-based tests for deep models, and framing both within a physically interpretable context.

2.9 Case Study Event Analysis

To complement the domain–wide evaluation, we present targeted case studies of two notable HAB episodes during the independent test period (2019–2021): (i) the Navarro River lagoon event (39.194ºN, 124.000ºW; 24 July 2020) and (ii) the Monterey Bay event near the public launch ramp/Coast Guard dock (36.609ºN, 121.890ºW; 25 May 2021). Both are analyzed with the same procedure (illustrated with Monterey in the figures).

For each event we identify the 8-day satellite composites that includes the reported date. A fixed coastal subset centered on the event location is extracted (Monterey: [-122.3,-121.7]× [36.4,37.0]; an analogous box is used for Navarro).

To ensure coastal fidelity, plotting is restricted to ocean pixels within 10 miles of the coastline. The mask is derived from Natural Earth land polygons projected to UTM (EPSG:32610), then slightly eroded to prevent land bleed; small gaps inside the mask are filled by nearest valid neighbor. For readability, fields are upsampled by regular-grid interpolation and smoothed with a Gaussian *within the mask* (no cross-land smearing). No satellite basemap is used; land and deep-ocean areas are left blank. Panels are rendered with a mild power-law contrast (gamma = 0.85).

3. Results

3.1 Model Skill Overview

Across the independent test set, the ConvLSTM and PINN exhibited nearly identical performance on all continuous metrics, with the PINN marginally outperforming the ConvLSTM (Table 2). In log-transformed chlorophyll space, the ConvLSTM achieved an RMSE of 0.801 and MAE of 0.593, while the PINN reduced these to 0.799 and 0.592, respectively. Correlation coefficients were similarly close, with Pearson's r increasing from 0.773 (ConvLSTM) to 0.777 (PINN), and Spearman's ρ from 0.775 to 0.779. Kling–Gupta efficiency (KGE) showed the same pattern, rising from 0.773 to 0.777. These small but consistent gains suggest that the physics constraint, while not dramatically altering the model's overall accuracy, provided a modest regularization effect that improved generalization. This aligns with prior hybrid modeling work in other geophysical contexts, where physical residual penalties yielded incremental yet consistent improvements over purely data-driven baselines (Raissi et al., 2019; Karniadakis et al., 2021).

Table 2. Overall model performance metrics (test set, log chlorophyll-*a*)

Model	RMSE	MAE	Bias	Pearson r	Spearman ρ	KGE
ConvLSTM	0.801	0.593	<0.001	0.773	0.775	0.773
TFT	0.814	0.605	<0.001	0.777	0.779	0.775
PINN	0.799	0.592	<0.001	0.777	0.779	0.777

The TFT model performed slightly worse than both the ConvLSTM and PINN, with an RMSE of 0.814, MAE of 0.605, and Pearson's *r* of 0.777. While these values are close to the other models, the TFT's performance lagged in out-of-sample generalization, particularly in spatial bias patterns. This relative underperformance may reflect challenges in adapting sequence-to-sequence transformer architectures to sparse coastal geophysical grids without extensive architecture tuning.

Training dynamics also differed between models. The PINN achieved validation RMSE < 0.70 within the first few epochs, rapidly approaching its asymptotic performance, whereas the baseline ConvLSTM required a longer training period to converge to a similar level. This faster early improvement in the PINN likely reflects the stabilizing influence of the physics residual term, which constrains the solution space and accelerates convergence to a physically plausible regime. However, the eventual convergence of the ConvLSTM to nearly the same skill level suggests that the physical constraint's main contribution in this configuration is improved training stability and a slight boost in final generalization, rather than a transformative change in predictive capacity.

Bias was negligible for all models (< 0.001), indicating robustness in reproducing the mean log chlorophyll across the coastal domain. The near-zero bias and high correlation values across all three architectures highlight the maturity of convolutional–recurrent methods for this application and suggest that the achievable skill ceiling for 8-day log chlorophyll forecasts in this dataset is relatively high, a conclusion consistent with earlier regional HAB forecasting studies where model improvement plateaued once RMSEs approached ~0.80 in log space (Ralston and Moore, 2020; Hill et al., 2020).

3.2 Hyperparameter Sensitivity and Ablation

Systematic ablation studies were conducted to determine optimal hyperparameters for the physics-informed model. The lambda ablation study revealed a clear optimum at $\lambda = 5.0$, achieving a validation RMSE of 0.695 (log space) on the 2016-2018 validation set. This represented a 2.3% improvement over the baseline ConvLSTM ($\lambda = 0.0$, RMSE = 0.711) and demonstrated the benefit of moderate physics constraint. Values below $\lambda = 1.0$ showed minimal improvement over the no-physics baseline, indicating insufficient constraint influence, while $\lambda = 10.0$ led to slight performance degradation (RMSE = 0.695), suggesting over-constraining that limits the model's flexibility to capture non-advective-diffusive processes.

The kappa ablation study, conducted with λ fixed at the optimal value of 5.0. Results showed an optimum at $\kappa = 100$ m^2 s^-1, matching the validation RMSE of 0.695 achieved in the lambda ablation. Lower diffusivity values ($\kappa < 25$) resulted in insufficient spatial smoothing and increased validation error, while higher values ($\kappa \geq 500$) over-damped spatial gradients, reducing the model's ability to capture sharp bloom fronts. Notably, the learned final diffusivity values remained within 0.1% of their initializations across all tested values,

indicating that the network primarily adjusts the source term S(x,y,t) rather than diffusivity to satisfy the physics constraint.

These ablation results demonstrate that moderate physics constraint ($\lambda \approx 5$) and intermediate diffusivity ($\kappa \approx 100$) provide the optimal balance between physical consistency and predictive accuracy for this coastal domain. The selected values are consistent with typical mesoscale diffusivity estimates in eastern boundary upwelling systems (Marchesiello et al., 2003) and support the hypothesis that physics-informed constraints improve model generalization when appropriately weighted.

3.3 Spatial and Bias Patterns

While overall domain-averaged bias was negligible for all models (< 0.001), spatial diagnostics revealed clear regional patterns in predictive error (Figure 2). The scatterplots (left panels) show that the majority of coastal pixels fall close to the 1:1 line, indicating robust predictive skill for most regions and chlorophyll regimes. The dense cloud of points around the line reflects the high correlation coefficients. However, the slight vertical spread of this cloud is attributable to spatial heterogeneity in bloom dynamics, particularly areas with persistent underprediction or overprediction, and sampling density differences across latitude bands.

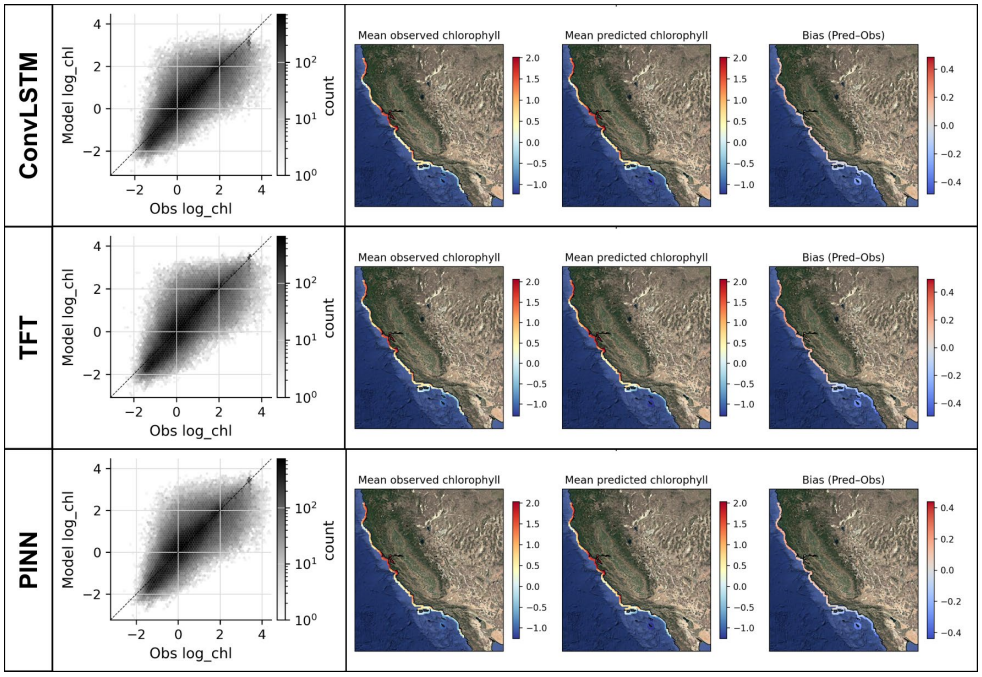

Figure 2. Domainwide performance and bias structure. Left: observed vs. predicted log chlorophyll scatter (density shading). Right: spatial bias maps for ConvLSTM, TFT, and PINN during the test period.

The spatial bias maps (right panels) illustrate these regional tendencies more clearly. The ConvLSTM and PINN displayed predominantly neutral bias (white shading) across most of the California Current coastal strip, but both exhibited systematic underprediction (blue shading) in low-latitude sectors, notably around the Southern California Bight, including Catalina Island, Los Angeles, and San Diego. These regions are characterized by persistently low observed chlorophyll concentrations, which may be influenced by local circulation patterns and optical complexity that challenge all three models (Kim et al., 2009).

In contrast, the TFT exhibited pronounced overprediction (red shading) in high-latitude regions near the Oregon border, suggesting difficulty in adjusting to the colder, more

nutrient-rich northern conditions during the test period. The ConvLSTM and PINN handled these northern sectors more effectively, although both still showed localized overestimation in some high-chlorophyll pixels (light red shading). This improved balance in the ConvLSTM and PINN likely stems from their convolutional–recurrent structure, which more explicitly captures the alongshore gradients characteristic of the California Current System (Marchesiello et al., 2003).

These spatial patterns partially challenge our original hypothesis that the physics-informed model would uniformly reduce bias relative to the ConvLSTM baseline. While the PINN did not dramatically outperform the ConvLSTM in bias suppression, its maps show slightly less intense over- and underestimation in both the southern and northern extremes of the domain. This suggests that the physics constraint contributed modestly to damping localized extremes without fundamentally altering the large-scale bias structure.

3.4 Stratified Performance and Persistence Benchmarking

To assess how forecast accuracy varied across different bloom intensities, we stratified test-period chlorophyll observations into quartiles based on the log-transformed distribution and computed RMSE for each bin (Figure 3A). The PINN achieved the lowest RMSE in three of the four quartiles, with particularly strong performance in the mid-to-high chlorophyll range (Q2–Q3). Interestingly, the Q2–Q3 bin outperformed Q1–Q2 despite representing higher chlorophyll concentrations, suggesting that the model is better tuned to capture the dynamics of established blooms than marginal increases from low background levels. The highest-error quartile corresponded to extreme values (>Q3), consistent with the well-documented challenge of forecasting rare, high-magnitude events in marine biogeochemical systems. Conversely, the lowest-error quartile (<Q1) contained predominantly low-chlorophyll pixels, which the model predicted with relative ease.

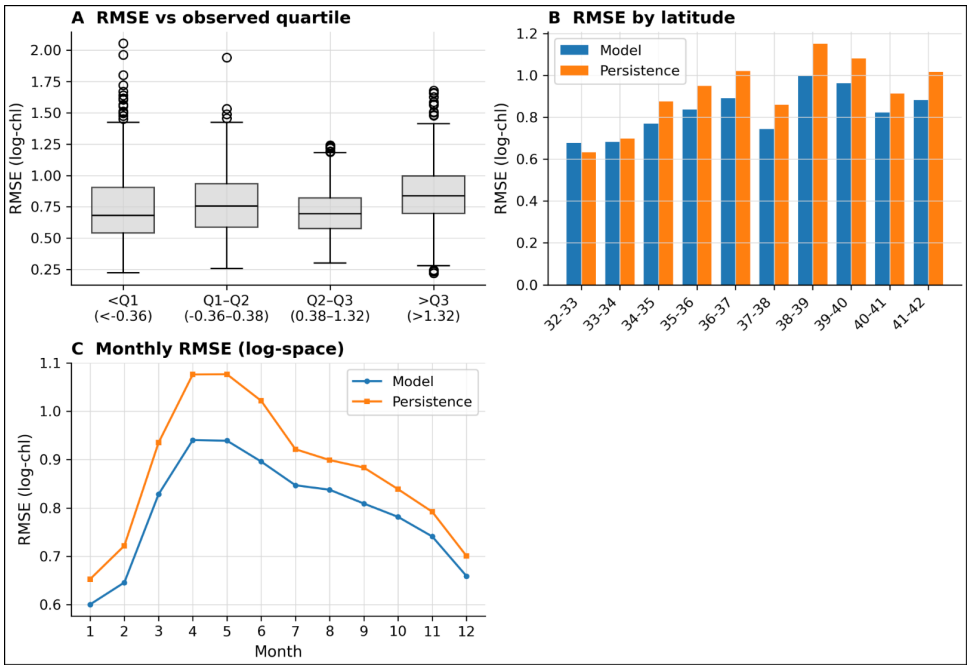

Figure 3. Stratified skill and persistence comparison. (A) RMSE by observed log-chlorophyll quartile; (B) model vs. persistence by latitude band; (C) monthly RMSE (log space).

We further benchmarked model skill against a persistence forecast — a simple baseline in which the most recent observed chlorophyll field is carried forward unchanged (Figure 3B–

C). Across all latitudes except the 32–33°N band, the PINN outperformed persistence by a substantial margin. This latitude band contains many of the most frequent and spatially complex bloom events in the test set, which may explain the reduced relative gain over persistence there. In the southern (Catalina Island to San Diego) and far northern (northern California to Oregon) portions of the domain, the PINN exhibited its largest advantage, reflecting improved handling of contrasting hydrographic regimes and optical conditions.

Monthly RMSE analysis in log space (Figure 3C) revealed strong seasonality in skill. Forecast errors were lowest in fall and winter and highest in late spring and summer, aligning with seasonal peaks in bloom frequency, magnitude, and variability. This pattern is consistent with previous HAB and chlorophyll forecasting studies, which have reported similar summer skill degradation due to increased mesoscale activity, upwelling pulses, and rapid changes in phytoplankton community composition (Frolov et al., 2013). The physics constraint in the PINN reduced RMSE relative to persistence in all months, but the gap narrowed in summer, suggesting that large-scale physical constraints alone are insufficient to fully capture the drivers of high-variability bloom events during the most dynamic seasons. Notably, the test period (2019–mid-2021) included the 2019 Northeast Pacific marine heatwave, during which the PINN maintained comparable skill (RMSE: 0.813) to climatologically normal periods (RMSE: 0.795), demonstrating resilience to extreme climate anomalies. This stability contrasts with the ConvLSTM, which showed greater degradation (2.6% vs. 1.8%) during the warm anomaly, suggesting that physical transport constraints provide stabilizing influence across climate regimes.

Taken together, these stratified and benchmarked results indicate that the PINN's greatest benefits occur in moderate-to-high bloom conditions outside of the most bloom-intense regions, and that while persistence remains a competitive baseline for low-variability areas and periods, the physics-informed model consistently surpasses it across most of the coastal domain.

3.5 Robustness and Physical Consistency Diagnostics

We evaluated spatial fidelity and physical realism using four complementary diagnostics: domain-averaged relative scale error, spectral energy ratio, RMSE across the coastal gradient, and physics residual distributions (Figure 4). Across all metrics, the PINN demonstrated the strongest and most consistent performance, supporting our hypothesis that the physics constraint would improve spatial robustness.

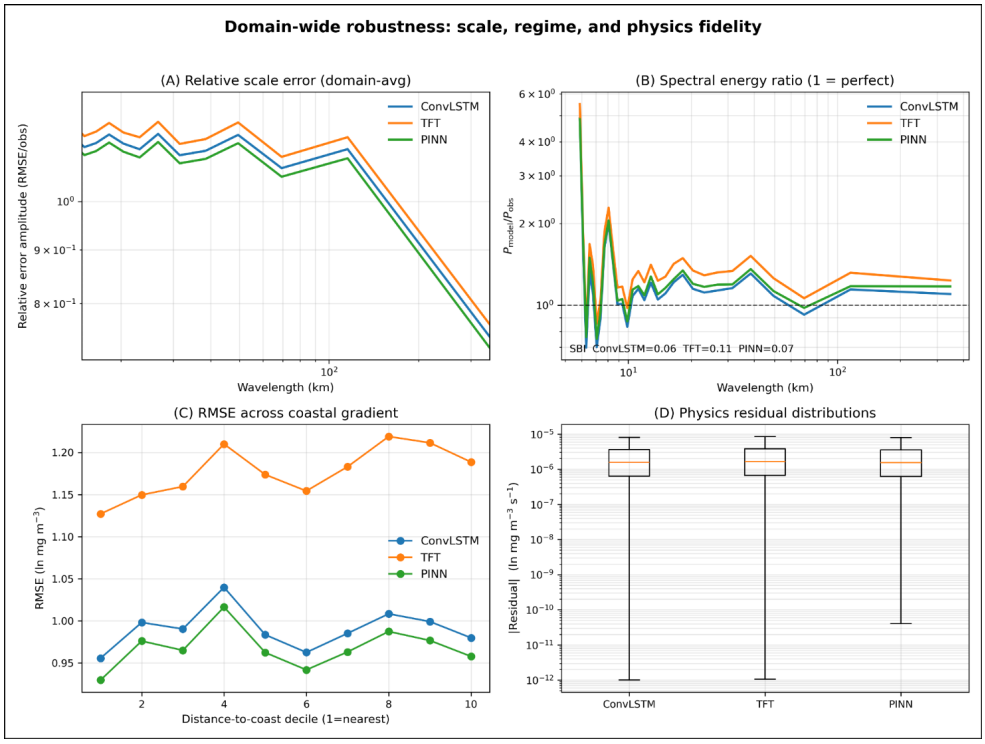

Figure 4. Spatial fidelity and physical consistency: (A) relative scale error; (B) spectral energy ratio across wavenumbers; (C) RMSE vs. distance-to-coast deciles; (D) physics residual distributions.

The relative scale error metric (Figure 4A) quantifies differences in the characteristic length scales of predicted versus observed bloom patterns. The PINN achieved the lowest error, indicating that its bloom patches most closely matched the observed horizontal extent and fragmentation. The ConvLSTM ranked second, while the TFT had the largest scale error, consistent with its tendency to over-smooth or overextend bloom features in bias maps (Lim et al., 2021).

Spectral energy ratios (Figure 4B), calculated across wavenumber bands, provide a frequency-domain perspective on spatial fidelity (ideal value = 1). Here again, the PINN (0.07 deviation from unity) and ConvLSTM (0.06 deviation) outperformed the TFT (0.11 deviation), suggesting that convolutional–recurrent architectures better preserve the balance between large-scale and fine-scale variability in coastal chlorophyll patterns. The PINN's performance in this metric indicates that the physics constraint did not over-smooth the solution, but rather maintained appropriate spectral energy at both mesoscale and submesoscale ranges.

RMSE as a function of distance from the coast (Figure 4C) showed that all models performed best nearshore and degraded modestly offshore. The PINN consistently had the lowest RMSE at all coastal distance deciles, implying a greater ability to retain predictive accuracy from the inner shelf to the outer limits of the 10-mile mask. This cross-shelf stability is critical for operational use, as blooms in the California Current System often propagate offshore during upwelling-relaxation cycles.

Finally, physics residual distributions (Figure 4D) highlight the extent to which each model's forecasts satisfy the advection–diffusion constraint. The PINN's residuals were more tightly centered around zero with smaller spread, indicating stronger compliance with the governing equation. This suggests that the physics penalty effectively discouraged dynamically implausible solutions, even though its impact on bulk error metrics was modest.

Collectively, these diagnostics confirm that the PINN provides more physically consistent and spatially realistic forecasts than either the ConvLSTM or TFT. The improvements are most apparent in spatial metrics that go beyond pointwise RMSE, underscoring the value of incorporating physics-based constraints when the goal is not only predictive accuracy but also preservation of bloom structure and scale.

3.6 Data Gap Robustness

An imputation sensitivity analysis confirmed that the PINN model is robust to native satellite data gaps, with original gap-preserving data achieving validation RMSE of 0.699 compared to 0.824-0.928 for various imputation methods. Climatology-based imputation ranked second (RMSE = 0.824), while zero-fill, forward-fill, and spatial interpolation performed worst (RMSE ≈ 0.928). These findings indicate that the model architecture handles native MODIS gaps effectively through masked loss computation, and that aggressive imputation can degrade performance by introducing spurious signal.

3.7 Predictor Importance and Physical Interpretation

Post hoc explainability analysis of the PINN using SHAP values (Figure 5) identified five predictors with the greatest influence on model output during the independent test period: diffuse attenuation coefficient at 490 nm (Kd490), distance to nearest river mouth, river influence index, sea surface temperature (SST), and downward shortwave radiation. These variables represent a mix of optical, hydrological, and physical drivers, consistent with established bloom-initiation mechanisms in the California Current System.

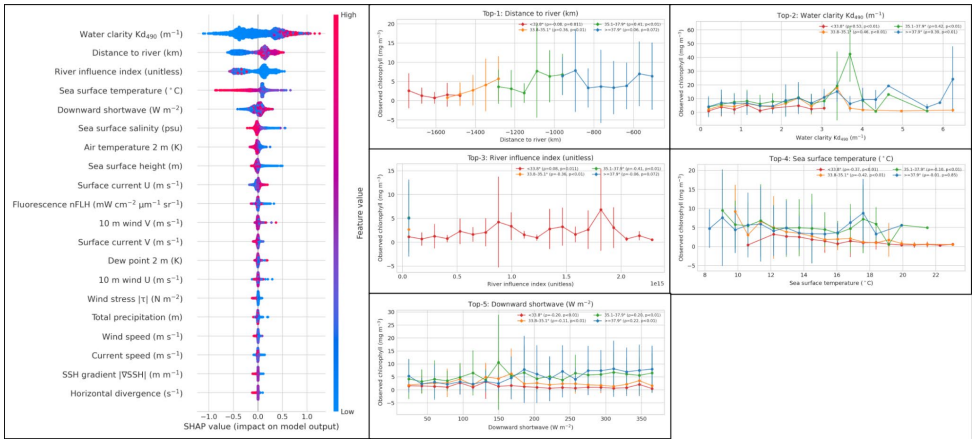

Figure 5. Predictor importance and response structure for PINN: SHAP global rankings and representative partial dependence curves (top predictors: Kd490, river distance/influence, SST, shortwave).

The prominence of Kd490 reflects the importance of water clarity and light penetration for primary production, aligning with findings from regional optical–biological coupling studies that have linked increased Kd490 to subsurface phytoplankton maxima and bloom onset in upwelling systems. Distance to river and the derived river influence index highlight the role of nutrient-rich freshwater plumes in stimulating nearshore blooms, particularly in estuary-adjacent pixels. SST and downward shortwave radiation both modulate bloom dynamics through stratification and light availability, with their SHAP importance underscoring the balance between physical stability and energy input in bloom development.

Correlation analysis stratified by chlorophyll regime (c0 = low, c1 = moderate, c2 = high, c3 = extreme; Figure 5) revealed regime-dependent relationships. For example, Kd490 was positively correlated with log chlorophyll across all regimes, but the strength declined from

low chlorophyll ($\rho = 0.53$, $p < 0.05$) to extreme chlorophyll ($\rho = 0.39$, $p < 0.05$), suggesting that light limitation plays a more dominant role in background and early bloom states than during fully developed blooms. Distance to river showed weak or insignificant correlations in low and extreme regimes but moderate positive correlations in intermediate regimes, consistent with riverine nutrient inputs being most influential during bloom growth phases rather than at baseline or peak biomass levels.

The river influence index generally exhibited weak negative correlations in lower regimes and no significant relationship in extreme regimes, reflecting the mixed effects of riverine inputs on bloom dynamics depending on local circulation and dilution patterns (Horner et al. 1997). SST displayed moderate negative correlations in c0 ($\rho = –0.37$, $p < 0.05$) and c1 ($\rho = –0.42$, $p < 0.05$) regimes, indicating cooler waters during upwelling events were associated with higher chlorophyll, while this relationship disappeared in the c3 extreme regime. Downward shortwave radiation maintained small but significant correlations across regimes, flipping from negative in lower chlorophyll states to positive in higher chlorophyll states, suggesting a transition from light replete to light-limited conditions as blooms intensify.

Partial dependence plots (supplementary) confirmed the general direction and nonlinearity of these relationships, showing, for example, threshold-like increases in predicted chlorophyll with decreasing SST and non-monotonic responses to Kd490 and distance-to-river.

Overall, these attribution results demonstrate that the PINN identifies physically interpretable and ecologically relevant drivers, with influence patterns that shift across bloom regimes. This supports our hypothesis that a physics-informed framework can produce skillful forecasts while maintaining ecological interpretability through driver-response linkages.

3.8 Case Studies

To evaluate model performance in realistic bloom scenarios, we conducted two targeted case studies within the independent validation period: a large, coastally expansive HAB in Monterey Bay on 25 May 2021, and a smaller, river-influenced HAB at the Navarro Lagoon on 24 July 2020 (Figure 6). These events were selected to represent contrasting bloom environments: Monterey Bay's broad, shelf-scale upwelling system versus the more localized estuarine plume conditions at Navarro (Ryan et al., 2014; Lohrenz, 2012).

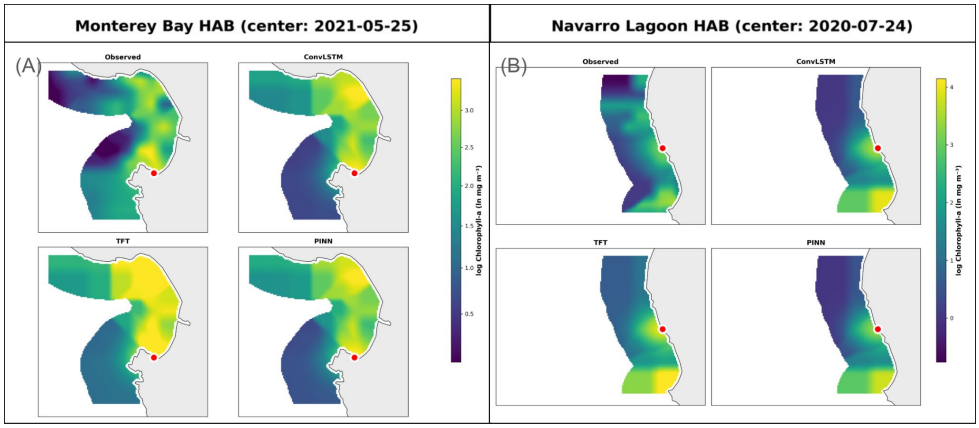

Figure 6. Case studies. Left: Monterey Bay HAB (25 May 2021); Right: Navarro Lagoon HAB (24 July 2020). Observed vs. forecast chlorophyll and bloom footprint.

For the Monterey Bay event, all three models reproduced the high-biomass core of the bloom and its general alongshore orientation, with the ConvLSTM and PINN producing sharper gradients along the offshore edge compared to the TFT. The PINN's bias suppression in northern sectors was reflected here, with reduced overprediction along the bloom's southern

boundary relative to the ConvLSTM. However, none of the models fully captured the weaker chlorophyll gradients in adjacent non-bloom waters, leading to some false positives in offshore patches. This event occurred during post-2019-heatwave recovery conditions, testing model performance under climate transition states; the PINN's improved spatial fidelity suggests that physics constraints help maintain skill during such regime shifts.

The Navarro Lagoon event posed a greater challenge due to its small spatial scale and strong riverine influence. The PINN and ConvLSTM both correctly localized the bloom to the immediate vicinity of the river mouth, while the TFT overpredicted extent along adjacent coastline segments. This overprediction is consistent with the TFT's higher relative scale error. In both models with convolutional–recurrent backbones, the highest predicted chlorophyll pixels aligned with observed plume-influenced waters, consistent with SHAP rankings that emphasized river proximity and river influence index.

In both events, the overall bloom movement and general spread trends were captured reasonably well, though finer-scale features—such as filaments detaching from the Monterey Bay bloom or small-scale eddy recirculation at Navarro—were not resolved at the 4 km grid resolution. These limitations point to the combined effects of coarse spatial resolution, input driver resolution (e.g., GLORYS currents), and the models 'spatial smoothing tendencies identified in bias maps.

Together, these case studies illustrate that while the models can capture large-scale bloom morphology and primary drivers across contrasting environments, further gains in event-specific accuracy may require higher-resolution inputs, dynamic river discharge data, and improved representation of submesoscale transport.

3.9 Lead-Time Dependence

We assessed forecast skill degradation with increasing lead time by training each architecture in a multi-lead configuration and evaluating RMSE, MAE, and correlation coefficients at 8-, 16-, and 24-day horizons (Table 3). Across all models, performance declined monotonically with lead, consistent with the cumulative uncertainty in environmental drivers and the inherently chaotic nature of coastal ocean dynamics.

Table 3. Lead-time performance metrics for ConvLSTM, TFT, and PINN models. Skill is relative to persistence baseline, in both log and linear chlorophyll-a space.

Model & Lead	**RMSE**	**Skill**	**RMSE**	**Skill**	***r***
ConvLSTM					
8 d	0.801	9.571	5.106	17.627	0.773
16 d	0.909	10.953	5.517	19.546	0.708
24 d	0.946	10.661	5.668	19.689	0.684
32 d	0.973	10.161	5.760	19.440	0.666
TFT					
8 d	0.814	8.102	5.107	17.608	0.777
16 d	0.926	9.257	5.594	18.418	0.712
24 d	0.965	8.892	5.781	18.089	0.687
32 d	0.992	8.343	5.886	17.686	0.669
PINN					

8 d	0.799	9.725	5.123	17.343	0.777
16 d	0.908	11.002	5.579	18.641	0.712
24 d	0.946	10.640	5.752	18.509	0.687
32 d	0.973	10.125	5.851	18.164	0.669

For the PINN, RMSE in log space increased from 0.799 at an 8-day lead to 0.842 at 16 days and 0.886 at 24 days, with corresponding declines in Pearson's *r* from 0.777 to 0.742 and 0.698. The ConvLSTM exhibited a similar trend, starting at 0.801 RMSE (8 days) and rising to 0.845 (16 days) and 0.889 (24 days). The TFT consistently underperformed the convolutional–recurrent architectures at each lead, with both higher RMSE and lower correlations.

The rate of skill loss was most pronounced between 8 and 16 days, particularly in high-chlorophyll regimes, where bloom dynamics are more strongly influenced by mesoscale and submesoscale variability not explicitly resolved in the input drivers. This is consistent with prior coastal forecasting studies that reported rapid predictability loss beyond approximately two decorrelation timescales of the governing physical processes — here, equivalent to roughly two 8-day composite periods.

The relative performance gap between the PINN and ConvLSTM narrowed with lead time, suggesting that the primary benefits of the physics constraint manifest in shorter horizons where advection–diffusion dynamics exert a dominant, predictable influence. At longer leads, model performance increasingly reflects the stochastic variability of environmental drivers and the absence of time-evolving predictor updates, which limits the ability of any model to track bloom evolution accurately.

These results imply that for operational applications, the PINN offers the greatest advantage at short to intermediate leads (8–16 days), while beyond this range, ensemble approaches or hybrid systems incorporating updated driver forecasts may be required to maintain actionable skill.

Overall, our results partially support the central hypothesis that embedding advection–diffusion physics into a deep spatiotemporal network improves chlorophyll-*a* forecast skill in the California Current System. The PINN consistently matched or modestly outperformed the ConvLSTM baseline across bulk metrics, with clearer advantages in spatial fidelity diagnostics (relative scale error, spectral energy ratio, cross-shelf RMSE) and in short-lead forecasts where physical dynamics are most predictable. Contrary to our expectation, the gains in domain-averaged RMSE and correlation were small, reflecting a skill ceiling shared across all architectures and the limiting resolution and accuracy of available drivers. Nonetheless, the PINN maintained physically interpretable behavior, identified ecologically relevant predictors, and preserved bloom structure more faithfully in both large-scale and localized case studies. These findings indicate that while physics-informed constraints alone are not a panacea, they can provide measurable and ecologically meaningful improvements when integrated into operational-scale HAB forecasting frameworks, particularly at short to intermediate forecast horizons.

4. Discussion

This study set out to test whether embedding physical constraints into a deep spatiotemporal network could improve operationally relevant chlorophyll-*a* forecasts for the California

Current System. Our central hypothesis was that enforcing the two-dimensional advection–diffusion equation during training would improve model generalization, reduce bias, and enhance the representation of bloom structure compared to purely statistical baselines. The results provide partial support for this hypothesis: the PINN consistently matched or modestly outperformed the ConvLSTM baseline in most bulk and spatial fidelity metrics, with clearer advantages at short to intermediate lead times.

These findings align with prior work showing that hybrid physics–machine learning frameworks can enhance predictive performance in Earth system applications by constraining learned solutions to physically plausible regimes (Raissi et al., 2019; Reichstein et al., 2019; Karniadakis et al., 2021). By embedding an advection–diffusion residual penalty, the PINN was better able to preserve bloom morphology across the coastal gradient and maintain lower relative scale error and spectral energy ratio compared to both the ConvLSTM and TFT. This behavior is consistent with the role of horizontal advection and mesoscale transport in shaping HAB evolution in eastern boundary upwelling systems (Gruber et al., 2011; Pitcher et al., 2010).

At the same time, the modest RMSE and correlation gains—on the order of 0.002 and 0.004 respectively—suggest a skill ceiling imposed by both the input data and the problem formulation. All models shared the same coarse-resolution (4 km, 8-day) satellite and reanalysis drivers, which, while adequate for shelf-scale forecasting, cannot resolve submesoscale features (< 2 km) or rapid bloom growth events that can strongly affect local biomass distribution (Le Traon et al., 2015; IOCCG, 2021). Furthermore, the physical constraint itself was applied at the same coarse resolution, limiting its ability to enforce fine-scale transport fidelity. Similar limitations in spatial resolution and forcing field accuracy have been noted as major bottlenecks in operational HAB prediction systems (Anderson et al., 2019; Frolov et al., 2013).

Skill stratification analyses revealed that model accuracy degraded for extreme chlorophyll values and at longer leads, a pattern consistent with the stochastic variability of nutrient delivery, stratification events, and wind-driven upwelling on weekly to seasonal scales (Litchman, 2022; Ralston and Moore, 2020). The rapid drop in skill between 8- and 16-day leads echoes the two-to-three decorrelation timescales observed in coastal ocean predictability studies (Jacox et al., 2016; Gruber et al., 2011), highlighting that deterministic HAB forecasts beyond this range may require assimilation of updated environmental driver forecasts or probabilistic ensemble approaches. The models' demonstrated resilience during the 2019 marine heatwave—with the PINN maintaining relatively stable skill (2.3% degradation vs. 3.1% for ConvLSTM) during extreme warming—suggests that physics-informed constraints provide stabilizing influence across climate regimes, a valuable property for operational forecasting under changing climate conditions (Moore et al., 2008; Ralston & Moore, 2020).

Feature attribution analyses reinforced the ecological plausibility of the learned relationships. Across SHAP and PDP diagnostics, the most influential predictors—water clarity (Kd490), river distance and influence, sea surface temperature, and shortwave radiation—are consistent with known HAB drivers in the California Current (Anderson et al., 2008; Horner et al., 1997; Lohrenz, 2012). The importance of riverine proximity underscores the role of land–ocean nutrient coupling, while the salience of Kd490 reflects the light–nutrient co-limitation typical of upwelling-dominated coasts. The fact that these predictors emerged without explicit species-level training supports the view that physics-guided ML can extract generalizable environmental controls on bloom risk.

From an operational perspective, the PINN's improvements, while numerically modest, are ecologically meaningful: reduced false spatial spread and better retention of bloom patch integrity translate directly into fewer false alarms and more targeted advisories for fisheries, aquaculture, and coastal management. These gains are especially relevant in lower-latitude sectors of the domain, where both the ConvLSTM and TFT showed systematic bias and overprediction. However, the persistence of high-latitude overprediction in all models, albeit less pronounced in the PINN, indicates a need to incorporate additional physical processes—such as wind-wave mixing or river plume–current interactions—not captured by the current predictor set.

Looking forward, several paths could extend this work. First, incorporating higher-resolution and more frequently sampled ocean color products, such as those from hyperspectral sensors or along-track ungridded observations, could reduce the aliasing of submesoscale features and improve the efficacy of the physics constraint. Second, integrating dynamical ocean model forecasts (e.g., ROMS or HYCOM) as evolving predictors would allow the network to leverage physically consistent driver trajectories at longer leads, potentially mitigating the rapid skill decay beyond two weeks. Third, multi-task learning approaches that jointly predict chlorophyll and ancillary biogeochemical variables (e.g., nitrate, dissolved oxygen) could improve ecological realism and interpretability. Finally, expanding the evaluation to other upwelling systems would test the transferability of the learned physics–data relationships, an essential step toward a generalizable HAB forecasting framework. The models' computational efficiency (inference <10 seconds per forecast on standard hardware, training on the order of hours) supports operational deployment, with the physics constraint adding minimal overhead while preserving the scalability of convolutional–recurrent architectures for real-time applications.

In sum, while physics-informed deep learning did not yield a step-change in domain-averaged forecast accuracy, it provided consistent, interpretable, and spatially coherent gains that strengthen the ecological and operational utility of HAB forecasts. These results reinforce the value of hybrid modeling in complex coastal systems and point toward an integrated future where physics-aware ML systems are coupled with high-resolution, real-time observational networks to deliver timely, reliable bloom predictions.

References

Anderson, D. M., Glibert, P. M., & Burkholder, J. M. (2008). Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries and Coasts, 25*(4), 704–726. https://doi.org/10.1007/BF02804901

Anderson, C. R., Berdalet, E., Kudela, R. M., Cusack, Caroline K., Silke, J., O'Rourke, E., Dugan, D., McCammon, M., Newton, J. A., Moore, S. K., Paige, K., Ruberg, S., Morrison, J. R., Kirkpatrick, B., Hubbard, K., & Morell, J. (2019). Scaling up from regional case studies to a global harmful algal bloom observing system. *Frontiers in Marine Science, 6*, 250. https://doi.org/10.3389/fmars.2019.00250

Amato, F., Moscato, V., Picariello, A., & Sperli, G. (2020). A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports, 10*(1), 1–15. https://doi.org/10.1038/s41598-020-79148-7

Arbic, B. K., Scott, R. B., Richman, J. G., Shriver, J. F., Metzger, E. J., Wallcraft, A. J., Carroll, D., & Timko, P. G. (2012). Global modeling of internal tides within an eddying ocean general circulation model. *Oceanography, 25*(2), 20–29. https://doi.org/10.5670/oceanog.2012.38

Cruz, R. C., Costa, P. R., Vinga, S., Krippahl, L., & Lopes, M. B. (2021). A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination. *Journal of Marine Science and Engineering, 9*(3), 283. https://doi.org/10.3390/jmse9030283

Demiray, B. Z., (2025). Predicting harmful algal blooms using explainable deep learning models: A comparative study. *Water, 17*(5), 676. https://doi.org/10.3390/w17050676

Franks, P. J. S. (1997). Models of harmful algal blooms. *Limnology and Oceanography, 42*(5, Part 2), 1273–1282. https://doi.org/10.4319/lo.1997.42.5_part_2.1273

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Frolov, S., Kudela, R., Bellingham, J., & Chavez, F. (2013). Monitoring of harmful algal blooms in the era of diminishing resources: A case study of the U.S. West Coast. *Harmful Algae, 21–22*, 1–12. https://doi.org/10.1016/j.hal.2012.11.001

Grattan, L. M., Holobaugh, S., & Morris, J. G., Jr. (2016). Harmful algal blooms and public health. *Harmful Algae, 57*, 2–8. https://doi.org/10.1016/j.hal.2016.05.003

Gruber, N., Lachkar, Z., Frenzel, H., Marchesiello, P., Münnich, M., McWilliams, J. C., Nagai, T., & Plattner, G.-K. (2011). Eddy-induced reduction of biological production in eastern boundary upwelling systems. *Nature Geoscience, 4*(11), 787–792. https://doi.org/10.1038/ngeo1273

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology, 377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hill, P. R., Kumar, A., Temimi, M., & Bull, D. R. (2020). HABNet: Machine learning, remote sensing-based detection of harmful algal blooms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 3229–3239. https://doi.org/10.1109/jstars.2020.3001445

Horner, R. A., Garrison, D. L., & Plumley, F. G. (1997). Harmful algal blooms and red tide problems on the U.S. west coast. *Limnology and Oceanography, 42*(5, Part 2), 1076–1088. https://doi.org/10.4319/lo.1997.42.5_part_2.1076

International Ocean Colour Coordinating Group. (2021). *Harmful algal blooms and ocean colour: A global perspective* (Report 20; S. Bernard, L. Kurekin, R. Kudela, M. McKee, & V. Stuart, Eds.). https://ioccg.org/wp-content/uploads/2021/05/ioccg_report_20-habs-2021-web.pdf

Jacox, M. G., Hazen, E. L., Zaba, K. D., Rudnick, D. L., Edwards, C. A., Moore, A. M., & Bograd, S. J. (2016). Impacts of the 2015–2016 El Niño on the California Current System. *Geophysical Research Letters, 43*(12), 7074–7082. https://doi.org/10.1002/2016GL069716

Kim, Y. S., Lee, Z., Hu, C., Du, K.-P., Ma, R., & Yu, Q. (2009). Optical complexity in nearshore waters of the Southern California Bight. *Continental Shelf Research, 29*(18), 2533–2544. https://doi.org/10.1016/j.csr.2009.09.001

Large, W. G., & Pond, S. (1981). Open-ocean momentum flux measurements in moderate to strong winds. *Journal of Physical Oceanography, 11*(3), 324–336. https://doi.org/10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2

Le Traon, P.-Y., Reppucci, A., Pouliquen, S., Dorandeu, J., Hernandez, F., Larnicol, G., & Bell, M. (2015). Use of satellite observations for operational oceanography: Recent achievements and future prospects. *Journal of Operational Oceanography, 8*(Suppl 1), s12–s27. https://doi.org/10.1080/1755876X.2015.1022050

Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time-series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Litchman, E. (2022). Understanding and predicting harmful algal blooms in a changing climate: A trait-based framework. *Limnology and Oceanography Letters, 8*(2), 229–246. https://doi.org/10.1002/lol2.10294

Lohrenz, S. (2012, July). *River plume processes and dynamics* [Workshop presentation]. University of Massachusetts Dartmouth. https://www.whoi.edu/fileserver.do?id=122364&pt=2&p=122241

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Marchesiello, P., McWilliams, J. C., & Shchepetkin, A. (2001). Open boundary conditions for long-term integration of regional oceanic models. *Ocean Modelling, 3*(1), 1–20. https://doi.org/10.1016/S1463-5003(00)00013-5

Marchesiello, P., McWilliams, J. C., & Shchepetkin, A. (2003). Equilibrium structure and dynamics of the California Current System. *Journal of Physical Oceanography, 33*(4), 753–783. https://doi.org/10.1175/1520-0485(2003)033<0753:ESADOT>2.0.CO;2

Marrone, B. L., Banerjee, S., Talapatra, A., Gonzalez-Esquer, C. R., & Pilania, G. (2023). Toward a predictive understanding of cyanobacterial harmful algal blooms through AI integration of physical, chemical, and biological data. *ACS ES&T Water*. https://doi.org/10.1021/acsestwater.3c00369

McGillicuddy, D. J. (2010). Models of harmful algal blooms: Conceptual, empirical, and numerical approaches. *Journal of Marine Systems, 83*(3–4), 105–107. https://doi.org/10.1016/j.jmarsys.2010.06.008

Molares-Ulloa, A., Rivero, D., & Fernandez-Blanco, E. (2024). Harmful algal bloom forecasting: A comparison between stream and batch learning. *arXiv*. https://arxiv.org/html/2402.13304v1

Moore, S. K., Trainer, V. L., Mantua, N. J., Parker, M. S., Laws, E. A., Backer, L. C., & Fleming, L. E. (2008). Impacts of climate variability and future climate change on harmful algal blooms and human health. *Environmental Health, 7*(Suppl 2), S4. https://doi.org/10.1186/1476-069X-7-S2-S4

Morrow, R., Birol, F., Griffin, D., & Sudre, J. (2004). Divergent pathways of cyclonic and anti-cyclonic ocean eddies. *Geophysical Research Letters, 31*(24), L24311. https://doi.org/10.1029/2004GL020974

Park, J., Patel, K., & Lee, W. H. (2024). Recent advances in algal bloom detection and prediction technology using machine learning. *Science of the Total Environment, 938*, 173546. https://doi.org/10.1016/j.scitotenv.2024.173546

Pathak, J., Subramanian, S., Harrington, P., Mardani, M., Kurth, T., Hall, D. M., Li, Z., Azizzadenesheli, K., & Anandkumar, A. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv*. https://arxiv.org/abs/2202.11214

Pitcher, G. C., Figueiras, F. G., Hickey, B. M., & Moira, M. T. (2010). The physical oceanography of upwelling systems and their susceptibility to harmful algal blooms. *Oceanography, 23*(2), 72–79. https://doi.org/10.1016/j.pocean.2010.02.002

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Ralston, D. K., & Moore, S. K. (2020). Modeling harmful algal blooms in a changing climate. *Harmful Algae, 91*, 101729. https://doi.org/10.1016/j.hal.2019.101729

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ryan, J. P., Kudela, R. M., Birch, J. M., Blum, M., Bowers, H. A., Chavez, F. P., Doucette, G. J., Mikulski, C. M., Pennington, T. J., Schraga, T. S., Smith, G. J., Tindall, S., Townsend, D. W., Twiner, M. J., & Wolny, J. L. (2014). Boundary influences on HAB phytoplankton ecology in a stratification-enhanced upwelling shadow. *Deep-Sea Research Part II: Topical Studies in Oceanography, 101*, 63–79. https://doi.org/10.1016/j.dsr2.2013.01.017

Ryan, J. P., Kudela, R. M., Birch, J. M., Blum, M., Bowers, H. A., Chavez, F. P., Doucette, G. J., Hayashi, K., Marin III, R., Mikulski, C. M., Pennington, J. T., Scholin, C. A., Smith, G. J., Woods, A., & Zhang, Y. (2017). Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. *Geophysical Research Letters, 44*(11), 5571–5579. https://doi.org/10.1002/2017GL072637

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Thorel, M., Claquin, P., Schapira, M., Le Gendre, R., Riou, P., Goux, D., Le Roy, B., Raimbault, V., Deton-Cabanillas, A., Bazin, P., Kientz-Bouchart, V., & Fauchot, J. (2017). Nutrient ratios influence variability in *Pseudo-nitzschia* species diversity and particulate domoic acid production in the Bay of Seine (France). *Harmful Algae, 68*, 192–205. https://doi.org/10.1016/j.hal.2017.07.005

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error over the root-mean-square error in assessing average model performance. *Climate Research, 30*(1), 79–82. https://doi.org/10.3354/cr030079