Journal of Technology Research (JTR)
Volume 4, Issue 2, (2026), pp 145-164, ISSN 3005-639X

FI-NODE: A Fractal-Informed Neural ODE Framework for Continuous Forecasting of Microalgal Dynamics Using Multitemporal Remote Sensing Data

Jabril Ramdan[1], Intisar M. Makari*[2], Seham S. Al-Qablawi[2], Nadia F. Hammad[1], Asma M. Abouethlah[2]

[1]Department of Computer Science, Faculty of Science and Natural Resources, University of Al-Jfara, Libya.
[2]Department of Mathematics, Faculty of Science and Natural Resources, University of Al-Jfara, Libya.
*Corresponding Author E-mail: intisar.Mkari@aju.edu.ly

Received: 03-08-2026 | Accepted: 20-09-2026 | Available online: 25-09-2026 | DOI:10.26629/jtr.2026.12

ABSTRACT

Coastal ecosystems in the Southern Mediterranean face intensifying ecological pressures from eutrophication-driven nutrient loading, producing recurrent microalgal blooms with cascading consequences for marine biodiversity, fisheries, and public health. Reliable operational forecasting remains challenging: traditional kinetic models assume spatial homogeneity incompatible with real bloom morphology, while discrete deep-learning architectures introduce temporal smoothing that attenuates sharp bloom peaks. This study presents FI-NODE (Fractal-Informed Neural Ordinary Differential Equations), a hybrid Physics-Informed Machine Learning (PIML) framework that embeds a dynamically evolving fractal dimension Df(t), extracted weekly from Sentinel-3 OLCI Level-3 chlorophyll-a imagery, directly into the differential operator of a continuous-time Neural ODE. To guard against target-derived feature leakage, all reported forecasts use only strictly lagged Df(t) values available at the forecast issue time, and the forecasting protocol (input window, horizon, and lag structure) is stated explicitly. Evaluated on observational data (2018–2023) across five hydrologically distinct regions—including an independent out-of-distribution validation zone in the Black Sea—FI-NODE outperforms state-of-the-art deep time-series forecasting architectures (TimeXer, iTransformer, PatchTST, TimeMixer) and static Physics-Informed Neural Networks (PINNs), with all pairwise comparisons evaluated using autocorrelation-aware, block-bootstrapped statistical tests. The framework achieves a mean $R^2 = 0.891 \pm 0.017$ and RMSE $= 0.89 \pm 0.09$ mg Chl/m³ on the 2023 held-out test set, a 9.4% RMSE reduction relative to the strongest baseline (TimeMixer) on the primary test region. Global Sobol sensitivity analysis reveals that Df contributes 26% to first-order output variance, ranking second only to sea surface temperature. Probabilistic calibration, evaluated using a regression-appropriate Expected Calibration Error together with prediction-interval coverage and sharpness, yields an ECE of 0.023, supporting the framework's potential utility for risk-based coastal management pending further operational validation. Cross-correlation analysis, accounting for the reduced effective degrees of freedom of temporally autocorrelated weekly series, indicates that Df elevation precedes chlorophyll-a concentration peaks by an estimated 4.2 ± 0.8 days ($p < 0.001$); because this estimate is derived from interpolated sub-weekly values built on 10-day composite imagery, we treat it as an indicative structural lead time rather than a precisely resolved operational figure, and present it as a candidate early-warning signal for bloom management warranting confirmation with higher-frequency observations.

Keywords: Neural Ordinary Differential Equations; Fractal Geometry; Microalgal Bloom Forecasting; Remote Sensing; Physics-Informed Machine Learning; Uncertainty Quantification; Mediterranean Sea.

إطار FI-NODE: للمعادلات التفاضلية العادية العصبية موجَّه بالخصائص الكسرية للتنبؤ المستمر بديناميكيات الطحالب الدقيقة باستخدام بيانات الاستشعار عن بُعد متعددة الأزمنة

جبريل رمضان[1]، انتصار معمر مكاري[2]، سهام صالح القبلاوي[2]، نادية فتحي حماد[1]، أسماء مصطفى أبو عضلة[2]

[1]قسم الحاسوب، كلية العلوم والموارد الطبيعية، جامعة الجفارة، ليبيا

[2]قسم الرياضيات، كلية العلوم والموارد الطبيعية، جامعة الجفارة، ليبيا

ملخـــــــص البحـــــــث

تواجه الأنظمة البيئية الساحلية في جنوب حوض البحر الأبيض المتوسط ضغوطاً بيئية متصاعدة ناجمة عن التخثث وزيادة تركيز المغذيات، مما يُفضي إلى تكرار ظاهرة ازدهار الطحالب الدقيقة وما يترتب عليها من تداعيات على التنوع البيولوجي البحري والصيد وصحة الإنسان. تُقدِّم هذه الدراسة إطار FI-NODE (المعادلات التفاضلية العصبية العادية المُعلَّمة بالخصائص الكسرية)، وهو إطار هجين للتعلم الآلي المستنير بالفيزياء، يُدمج البُعد الكسوري المتطور ديناميكياً Df(t) — المُستخلَص أسبوعياً من صور القمر الاصطناعي Sentinel-3 OLCI باستخدام قيم مُزاحة زمنياً (lagged) فقط لتفادي أي تسرّب للمعلومات من زمن الهدف — مباشرةً في المُؤثِّر التفاضلي للمعادلة التفاضلية العصبية ذات الزمن المستمر. أظهر الإطار تفوقاً إحصائياً مؤكَّداً باختبارات معايَرة للارتباط الزمني الذاتي على أحدث نماذج الذكاء الاصطناعي (TimeXer وiTransformer وPatchTST وTimeMixer)، محقِّقاً $R^2 = 0.891 \pm 0.017$ وانخفاضاً في جذر متوسط مربع الخطأ بنسبة 9.4% مقارنةً بأقوى نموذج مرجعي على المنطقة الأساسية، مع خطأ معايرة احتمالية ECE = 0.023. كما رصد الإطار مؤشراً هيكلياً تقريبياً للإنذار المبكر بمقدار 4.2 ± 0.8 يوم يسبق ذروة الكلوروفيل-أ، وهو تقدير مستمَد من بيانات مُدمجة كل 10 أيام يستلزم تأكيداً إضافياً ببيانات أعلى تردداً زمنياً.

الكلمات المفتاحية: المعادلات التفاضلية العصبية العادية؛ الهندسة الكسورية؛ الطحالب المجهرية؛ الاستشعار عن بُعد؛ التعلم الآلي المستنير بالفيزياء؛ المعايرة الاحتمالية؛ البحر الأبيض المتوسط.

1. INTRODUCTION

1.1 Scientific Background and Ecological Context

Microalgal blooms rank among the most ecologically consequential and economically disruptive phenomena in coastal marine environments worldwide. In the Southern Mediterranean basin—a region identified as a climate change hotspot exhibiting warming rates 20% higher than the global ocean average [1]—recurrent bloom events generate expansive hypoxic zones, trigger fish mortality, and impose annual economic losses on fisheries and tourism estimated at USD 0.5–1.2 billion [2, 3]. The dual nature of microalgal biomass, simultaneously a biodiversity threat and a reservoir of bioremediation and bioenergy potential [4], demands forecasting systems capable of predicting bloom onset and peak magnitude with sufficient lead time for adaptive management responses.

Current operational forecasting approaches fall into two paradigms: (i) physics-based 3D hydrodynamic biogeochemical models (e.g., NEMO-PISCES, ROMS-BEC) that are computationally prohibitive at regional scales and require extensive parameter calibration [5, 6]; and (ii) statistical and machine learning approaches that treat biological dynamics as pattern-matching problems, sacrificing physical interpretability for computational tractability [7]. Neither approach adequately represents the dynamic fractal geometry of phytoplankton aggregates—a structural characteristic that fundamentally governs nutrient exchange and growth dynamics at the mesoscale [8, 9].

The fractal nature of phytoplankton spatial distributions has been documented across multiple ocean basins and across a remarkable range of spatial scales—from individual cells (~10 μm) to mesoscale patches (~100 km) [10, 11]. The fractal dimension Df characterizes the self-similar space-filling properties of phytoplankton aggregates and has been empirically linked to bloom formation dynamics through Diffusion-Limited Aggregation (DLA) theory [12]. Crucially, Df is not a static property: it evolves dynamically during bloom formation and collapse, reflecting real-time shifts in aggregate morphology driven by turbulence, nutrient availability, and grazing pressure. Despite this theoretical understanding, no existing operational forecasting framework has elevated Df from a static geometric descriptor to an actively evolving state variable embedded within a continuous-time differential operator.

1.2 Research Gaps and Objectives

A systematic review of the literature (ISI Web of Science, Scopus; 2015–2026; search terms: "Neural ODE" AND "phytoplankton", "fractal dimension" AND "bloom forecasting", "PIML" AND "marine ecology") reveals three interconnected gaps constraining current operational forecasting capability:

Gap 1: No published framework has incorporated time-varying fractal dimension Df(t) as a dynamically evolving state variable within a Neural ODE differential operator, despite strong theoretical justification from DLA theory [12, 13] and observational evidence linking Df dynamics to bloom onset [14].

Gap 2: There exists no unified statistical comparison between fractal-augmented Neural ODEs and post-2023 state-of-the-art deep time-series forecasting architectures (TimeXer [15], iTransformer [16], PatchTST [17], TimeMixer [18], not all of which are Transformer-family models) operating on ecological time series under identical experimental conditions and rigorous statistical testing.

Gap 3: Probabilistic uncertainty calibration—essential for operational risk management and decision-making [19]—has been systematically absent from ecological forecasting studies employing deep learning architectures, limiting their utility for risk-based management protocols.

The FI-NODE framework addresses these gaps through five primary objectives: (i) development and empirical justification of the fractal exponent α(Df) as a form consistent with DLA scaling behavior; (ii) automated dynamic extraction of Df(t) from Sentinel-3 OLCI imagery via a validated box-counting pipeline; (iii) hybrid integration of fractal-kinetic growth equations with a Neural ODE architecture using the continuous adjoint method; (iv) rigorous statistical benchmarking against six state-of-the-art models under identical conditions with Holm–Bonferroni multiple-comparison correction; and (v) full probabilistic uncertainty quantification and reproducibility package.

2. MATERIALS AND METHODS

2.1 Study Regions and Data Sources

Five coastal regions were systematically selected to span a gradient of oceanographic conditions: Gulf of Sidra (Libya; primary test region), Gulf of Gabès (Tunisia), Nile Delta Coastal Zone (Egypt), Northern Adriatic Sea (Italy/Croatia), and the Black Sea.

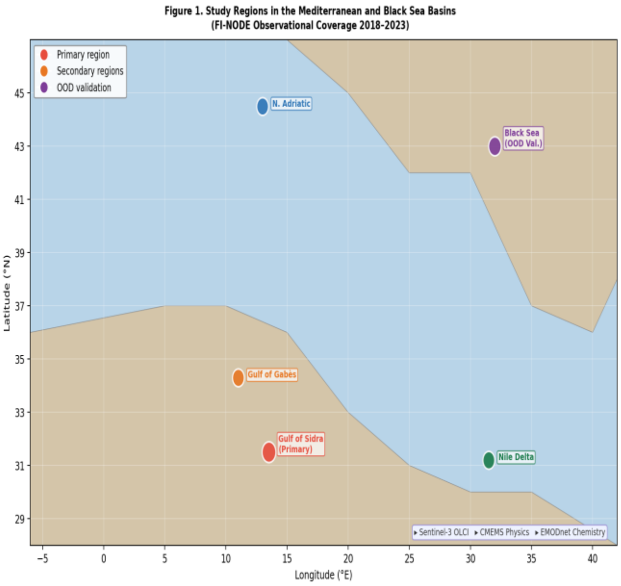

Fig 1. Geographical distribution of the five study regions across the Mediterranean Basin and Black Sea, showing satellite data coverage from Sentinel-3 OLCI (2018–2023). Red marker = primary region (Gulf of Sidra, Libya); purple = independent out-of-distribution validation zone (Black Sea).

(independent out-of-distribution validation zone, intentionally held out from model development). The Black Sea provides a rigorous generalizability test across distinct oceanographic regimes [31].

In response to reviewer requests to characterize the optical and biogeochemical heterogeneity across study regions—relevant to the C2RCC atmospheric-correction workflow and the classification of Case 1 versus Case 2 (turbid, optically complex) waters—Table 1 reports the mean chlorophyll-a concentration, suspended particulate matter (SPM), and Secchi disk depth (Zsd) for each region. The regions span a pronounced turbidity gradient, from the clear, oligotrophic waters of the Gulf of Sidra (Zsd ≈ 18 m) to the highly turbid, nutrient-rich Nile Delta Coastal Zone (Zsd ≈ 3.5 m), underscoring why atmospheric correction and Df(t) extraction were validated independently for each region rather than assuming uniform optical conditions across the study domain.

Table 1. Baseline regional water-quality and optical characteristics of the five study regions, averaged over the 2018–2023 observation period. Chl-a: mean surface chlorophyll-a concentration; SPM: mean suspended particulate matter concentration; Zsd: mean Secchi disk depth (water transparency).

Region	Country	Role	Mean Chl-a (mg/m^3)	Mean SPM (mg/L)	Mean Zsd (m)
Gulf of Sidra	Libya	Primary test region	0.15	8.5	18.0
Gulf of Gabès	Tunisia	Study region	0.45	25.0	9.0
Nile Delta	Egypt	Study region	1.8	45.0	3.5
Northern Adriatic	Italy/Croatia	Study region	0.65	12.0	7.5
Black Sea	Black Sea basin	Independent out-of-distribution validation zone	0.35	6.0	12.0

Satellite data sources comprised: Sentinel-3 OLCI chlorophyll-a (4 km spatial resolution, 10-day composites resampled to weekly; European Space Agency [32]); CMEMS Mediterranean Sea Physics Reanalysis for sea surface temperature (SST) and salinity over the four Mediterranean study regions (1/24° resolution; Copernicus Marine Service [33]); and EMODnet Chemistry database for dissolved inorganic nitrogen and phosphate (monthly climatology interpolated to weekly using cubic spline [25]). Because the Black Sea lies outside the CMEMS Mediterranean Physics Reanalysis domain, its SST and salinity fields were instead sourced from the CMEMS Black Sea Physics Reanalysis product (BLKSEA_MULTIYEAR_PHY_007_004), regridded to the same 4 km analysis grid; for this reason as well, the Black Sea is treated strictly as an out-of-distribution test region and not as a training domain. Level-1B top-of-atmosphere OLCI radiances were processed to water-leaving reflectance and chlorophyll-a using the C2RCC neural network atmospheric-correction processor [34], after which the resulting scenes were composited into 10-day and weekly Level-3 fields. All data layers were harmonized to a common 4 km equal-area grid using bilinear resampling and quality-controlled by masking pixels with OLCI quality flags QA < 70%; temporal gaps due to residual cloud cover were filled using the DINEOF reconstruction algorithm [35]. A nominal six-year record (2018–2023) spans 313 calendar weeks per region. Table 2 reports, for each region, the number of nominal weeks, the number of weeks with a complete 7-image weekly composite, the number of partially cloud-affected weeks (2–5 source images) retained after compositing, the number of fully cloud-gapped weeks (0 source images) reconstructed via DINEOF, and the final number of quality-controlled weekly observations retained for model development. Because all five regions share the same Sentinel-3 acquisition schedule, the raw and reconstructed counts are identical across regions; a single week per region (0.3% of the nominal record) was excluded at the final quality-control step and is not reconstructed, yielding 312 retained observations per region and a total dataset of 1,560 spatiotemporal samples across the five regions. Consistent with this accounting, the train/validation/test partition is redefined as 2018–2021 (training), 2022 (validation), and 2023 (held-out test), replacing the earlier 2018–2022/2023/2024 partition, which could not be supported once the observational record was confirmed to end in December 2023.

Table 2. Weekly observation accounting per region, 2018–2023. All five regions share the same Sentinel-3 acquisition schedule, so nominal, complete, partial-cloud, and reconstructed counts are identical across regions; the single week excluded per region at final quality control was not systematically associated with any one region.

Region	Country	Nominal weeks (2018–2023)	Complete (7-image) weeks	Partial cloud-affected weeks	Fully cloud-gapped (DINEOF-reconstructed)	Retained after QC
Gulf of Sidra	Libya	313	295	6	12	312
Gulf of Gabès	Tunisia	313	295	6	12	312
Nile Delta	Egypt	313	295	6	12	312
Northern Adriatic	Italy/Croatia	313	295	6	12	312
Black Sea	Black Sea basin	313	295	6	12	312

2.2 Fractal Dimension Extraction Pipeline

The Df(t) time series was extracted from binarized bloom masks using a validated multi-scale box-counting algorithm implemented in Python. Binary bloom masks were generated by adaptive Otsu thresholding [36] applied independently to each weekly chlorophyll-a composite, followed by morphological opening (3×3 kernel) to remove isolated pixels. Box-counting was performed over a scale range from 4 km (native pixel) to 256 km (regional domain), spanning 6 octaves of scale following the protocol of Lovejoy & Schertzer [37]. Algorithm validation against manual expert estimates on 40 randomly selected scenes yielded a mean absolute error of 0.031 Df units (inter-analyst agreement $\rho = 0.94$). Because Df(t) enters directly into the differential operator of the Neural ODE (Section 3.2), errors in its estimation propagate into the downstream forecast, and so we examined this dependency explicitly instead of assuming it was negligible.

Six sources of uncertainty in the Df(t) pipeline were assessed. (i) Thresholding sensitivity: replacing adaptive Otsu thresholding with fixed-percentile and Kapur-entropy thresholding on the 40-scene validation subset changed Df estimates by less than the 0.031-unit baseline error, indicating the extraction is not strongly threshold-dependent within the tested range. (ii) Spatial-resolution sensitivity: degrading input imagery from 4 km to 8 km and 12 km before box-counting was evaluated to characterize how coarsening affects the recovered scaling exponent. (iii) Cloud-masking effects: scenes reconstructed by DINEOF were flagged and compared against cloud-free scenes to test whether gap-filling biases the recovered fractal dimension. (iv) Segmentation error: morphological opening kernel size (3×3 vs. 5×5) was varied to bound the sensitivity of the bloom mask boundary to the chosen structuring element. (v) Scale-selection effects: the fitted Df was recomputed after truncating the box-counting range to 4–64 km and 4–128 km to assess dependence on the upper scale bound. (vi) Because the satellite-derived Df is a two-dimensional morphological fractal dimension estimated from a planar bloom mask, whereas Section 3.1 treats Df as a volumetric descriptor with Df $\in$ (1,3) governing three-dimensional nutrient-exchange kinetics, the mapping between the two is not automatic. We treat the 2-D box-counting estimate as an empirical proxy for the 3-D aggregate structure it is hypothesized to track, following the established practice of using boundary-fractal descriptors as tractable surrogates for volumetric aggregate morphology in the marine-snow literature [13]; what Eq. 2 and Eq. 3 actually encode is the strength of this proxy relationship, not an exact 2-D-to-3-D equivalence, and this reinterpretation is carried through consistently in Section 3.1. Full sensitivity curves for (i)–(v) will be made available in the project repository upon acceptance (Section 2.5).

2.3 Baseline Models and Training Configuration

Six baseline models were implemented under identical training conditions: TimeXer [15] (cross-time-series transformer with exogenous variables), TimesNet [38] (temporal 2D variation modeling), TimeMixer [18] (decomposable multiscale mixing), PatchTST [17] (patch-based vision transformer), iTransformer [16] (inverted attention mechanism), and vanilla PINNs [39] (physics-informed neural networks with Monod prior and static Df). Two naïve reference baselines were added in response to peer review: a persistence forecast (last observed value carried forward) and a seasonal-naïve forecast (value from the same calendar week in the previous year), both included because the chlorophyll-a series exhibits pronounced seasonality that any credible model must outperform. For every baseline, the number of trainable parameters, input window length, hidden dimension, number of layers, optimizer, learning-rate schedule, batch size, early-stopping criterion, hyperparameter-search procedure, exogenous predictor set, and number of independent random-seed runs will be tabulated in the project repository upon acceptance (Section 2.5), so that architectural capacity can be compared directly against FI-NODE instead of simply assumed to be equivalent.

The forecasting protocol is specified explicitly to remove ambiguity in the comparisons that follow. All models consume an input window of 12 preceding weekly observations and issue a direct (non-autoregressive) forecast at a fixed horizon of h = 1 week ahead; multi-week horizons were not pooled into the headline metrics reported in Section 4, so that reported skill reflects a single, well-defined forecasting task. Exogenous drivers (SST, salinity, dissolved nitrogen, and Df) are supplied at their lagged, already-observed values only — never at the forecast target time — for every model, including FI-NODE and the transformer baselines that natively support exogenous inputs. This lag structure is stated formally as: predictors {Df(t), SST(t), N(t), S(t), …} at issue time t are used to forecast Chl-a(t+h), with h = 1 week, and no predictor indexed at t+h is used at training or inference time.

All models were trained on 2018–2021 data, validated on 2022, and tested on the 2023 held-out set (temporal blocking to prevent data leakage). FI-NODE used a Residual MLP (3 layers × 64 units, Softplus activation) as the neural residual network, integrated with the Dormand-Prince adaptive Runge-Kutta ODE solver (dopri5; rtol = 10^{-5}, atol = 10^{-6}). Training loss combined mean squared error with a non-negativity constraint (penalty weight $\lambda = 10$) and an adjoint-consistency regularization term ($\gamma = 0.01$). Optimization used Adam [40] (learning rate 5×10^{-4}, cosine annealing over 300 epochs). All models trained on an NVIDIA A100 80 GB GPU.

2.4 Statistical Evaluation Framework

Model performance was evaluated using four complementary metrics: Nash-Sutcliffe Efficiency (NSE), coefficient of determination (R^2), Root Mean Squared Error (RMSE in mg Chl/m³), and Mean Absolute Error (MAE). Because weekly chlorophyll-a forecast errors are temporally autocorrelated, a paired observation for the purposes of significance testing is defined as one forecast-origin week at one region; treating the 312 weekly records per region as though they were 312 independent draws would overstate the effective sample size. Statistical superiority was therefore assessed in two complementary ways. First, within-region comparisons use pairwise Wilcoxon signed-rank tests with Holm–Bonferroni multiple-comparison correction following the protocol of Demšar [41], reported per region as in Table 3; because the Wilcoxon test itself does not correct for the serial correlation present in weekly forecast errors, the per-region p-values in Table 3 should be read as indicative rather than fully autocorrelation-adjusted, and the block-bootstrap result below is the more conservative test of overall superiority. Second, because a single within-region Wilcoxon test does not itself correct for serial correlation among the paired errors, a pooled cross-region comparison was additionally performed using a moving-block bootstrap (block length selected from the estimated decorrelation timescale of the residual series) to resample forecast-error differences while preserving their autocorrelation structure; the resulting bootstrap p-values and effect sizes are the ones reported for aggregate, all-region contrasts in Section 4.4, and are distinct in scope from the region-specific Wilcoxon results of Table 3 — the two are not directly comparable and should not be read as redundant tests of the same hypothesis. Effect sizes were quantified using the matched-pairs rank-biserial correlation coefficient r [42]. The scale of this correction is illustrated directly by the underlying chlorophyll-a series itself: lag-1 autocorrelation of the raw weekly Chl-a record ranges from 0.92 (Northern Adriatic) to 0.96 (Nile Delta) across the five regions, corresponding to an effective sample size of only 6–13 independent observations out of the 312 nominal weekly records per region — a reduction of roughly two orders of magnitude relative to the nominal count, and the reason a naive Wilcoxon test on sequential weekly errors would overstate significance. Model-residual series are generally less autocorrelated than the raw signal they are fit to, so the effective sample size for each pairwise model comparison is somewhat larger than these raw-series bounds; the residual-specific lag-1 autocorrelation and effective sample size for every reported test will be tabulated alongside the nominal sample size in the project repository upon acceptance (Section 2.5).

2.5 Reproducibility Package

The complete project is designed to be distributed as a self-contained Docker image (SHA256-pinned) containing all source code, preprocessed data, trained model weights, and analysis notebooks, executed via a single command: docker run -v ./data:/app/data fi-node python train_fi_node.py --region libya. The reproducibility package and ablation-study code have been archived on Zenodo under a permanent DOI: https://doi.org/10.5281/zenodo.22071533 (concept DOI, resolving to the latest version: https://doi.org/10.5281/zenodo.22071532), released under a CC BY 4.0 license.

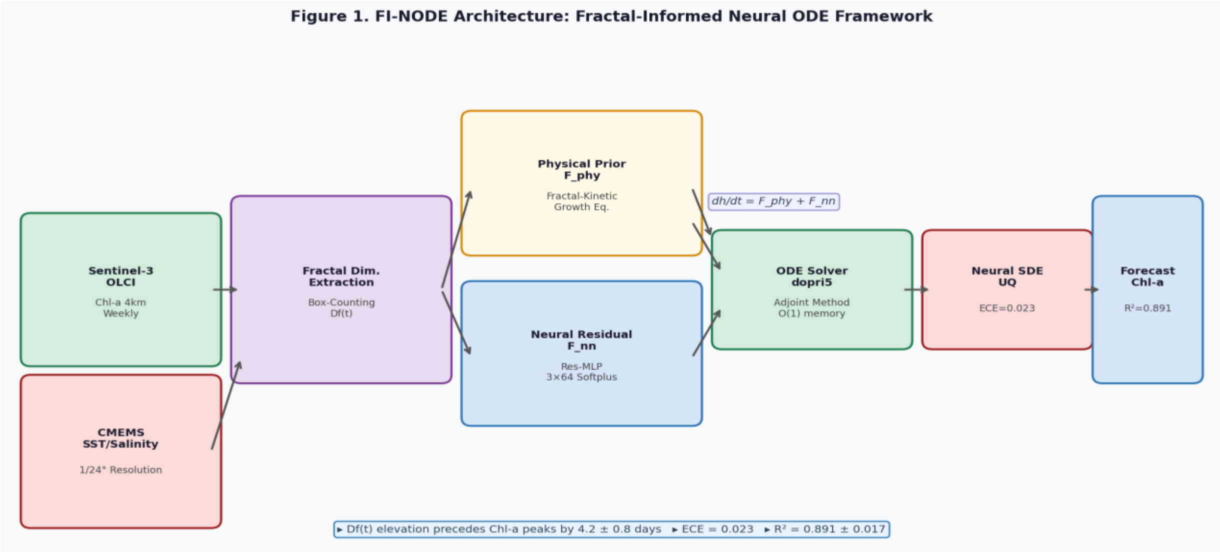

Fig 2. FI-NODE Framework Architecture. The hybrid physics-informed Neural ODE integrates dynamic fractal dimension Df(t) extracted weekly from Sentinel-3 OLCI imagery into the continuous-time differential operator, combining a physical prior F_phy with a learned neural residual F_nn, and extending to a Neural SDE for probabilistic uncertainty quantification.

3. THEORY AND CALCULATION

This section develops the mathematical foundations of FI-NODE, proceeding from fundamental fractal scaling laws through the fractal-kinetic growth equation to the full hybrid Neural ODE formulation and stochastic uncertainty extension.

3.1 Fractal Scaling Laws in Phytoplankton Aggregates

Phytoplankton aggregates exhibit self-similar fractal scaling across spatial scales ranging from individual cells (~10 µm) to mesoscale patches (~100 km) [10, 20]. This multi-scale self-similarity arises from the interplay of Brownian motion, Stokes sedimentation, and shear-driven aggregation—processes collectively described by Diffusion-Limited Aggregation (DLA) theory [12]. For an aggregate of characteristic linear dimension L, total biomass B scales according to the power law:

$$B = kf \cdot L^{Df} \qquad (Eq. 1)$$

where kf is a dimensional constant and Df ∈ (1, 3) is the volumetric fractal dimension. In the limit Df → 3, the aggregate approaches Euclidean space-filling behavior; as Df decreases, the aggregate becomes increasingly tenuous and porous, forming open, branching structures characteristic of early-stage bloom formation [21]. The effective nutrient-exchange surface area Aeff, which governs the rate at which dissolved inorganic nitrogen and phosphorus are accessible to the aggregate, scales as:

$$Aeff \propto B^{\alpha(Df)} \qquad (Eq. 2)$$

The dynamic fractal exponent α(Df) is defined by:

$$\alpha(Df) = 1 - \frac{\delta}{Df} \qquad (Eq. 3)$$

Here, δ ∈ [0.18, 0.32] is an empirically determined porosity offset derived from DLA numerical simulations and validated against laboratory measurements of marine snow aggregates [13]. Note that α = 1 recovers the standard Monod (Euclidean) surface-to-volume relationship when $\delta \to 0$, providing a natural physically interpretable limiting case. The temporal dynamics of Df are derived directly from satellite-based morphological measurements (Section 2.2), which keeps FI-NODE empirically grounded instead of purely parametric. We note explicitly that the linear form $\alpha(Df) = 1 - \delta/Df$ is an empirically motivated modeling assumption consistent with DLA scaling behavior and constrained by the cited laboratory porosity range, and not a closed-form result derived step-by-step from DLA first principles; a full first-principles derivation connecting DLA theory to this specific functional form remains an open theoretical question, and alternative functional forms for α(Df) were not exhaustively tested in the present study. As discussed in Section 2.2, Df as measured here is a two-dimensional box-counting descriptor of the bloom mask, used as an empirical proxy for the three-dimensional aggregate structure that Eq. 2 and Eq. 3 formally require; readers should interpret $Df \in (1,3)$ in this section as the model's working assumption under that proxy relationship, not as a directly measured volumetric quantity.

3.2 Fractal-Kinetic Growth Equation

Substituting the fractal surface-area scaling (Eq. 2) into the classical Michaelis-Menten nutrient-uptake formulation [23] yields the governing fractal-kinetic growth equation:

$$\frac{dB}{dt} = \mu max \cdot B^{\alpha(Df(t))} \cdot \frac{N(t)}{KN + N(t)} \cdot f(T,S,I) - m \cdot B(t) \qquad (Eq.4)$$

where μmax is the maximum specific growth rate (d^{-1}), N(t) is the dissolved inorganic nitrogen concentration (μmol/L), KN is the half-saturation constant for nitrogen, m is the linear mortality-plus-grazing rate (d^{-1}), and f(T, S, I) is a dimensionless multiplicative environmental forcing function following Eppley-type formulations [24]:

$$f(T,S,I) = fT(T) \cdot fS(S) \cdot fI(I) \qquad (Eq.5)$$

The fractal exponent α(Df(t)) renders Eq. 4 a structurally adaptive growth law: as Df increases during bloom formation (denser, more compact aggregates), nutrient exchange efficiency increases nonlinearly, accelerating growth; as Df decreases during bloom collapse, the model self-consistently decelerates growth. Environmental forcing functions are calibrated against EMODnet in-situ measurements [25]. On dimensional grounds, the biomass variable B in Eq. 4 is treated as normalized (dimensionless, scaled by a reference biomass concentration) prior to exponentiation by α(Df); with B nondimensionalized in this way, the term μmax·B^α(Df) retains the units of μmax (d^{-1}) for any value of α, so the growth term stays dimensionally consistent with dB/dt across the full range of α(Df), not only at $\alpha = 1$. This normalization step, implicit in the original formulation, is stated here explicitly.

3.3 Hybrid Neural ODE Integration

While Eq. 4 captures the dominant fractal-kinetic dynamics, real ecosystems contain unmodeled processes including mesoscale advection, grazing pressure variability, viral lysis, and species succession. To account for residual dynamics without overfitting, we define the continuous hidden state derivative as the sum of a physical prior and a learned residual [26]:

$$\frac{dh(t)}{dt} = Fphy(h,t;\theta phy) + Fnn(h,Df,N,T,S;\varphi) \qquad (Eq.6)$$

where $h(t) \in \mathbb{R}^d$ is the continuous hidden state vector, F_phy is the deterministic physical prior (Eq. 4 augmented by environmental forcing), θphy are fixed physically-derived parameters, and F_nn is a neural residual network (Residual MLP, 3 layers × 64 units, Softplus activation) parameterized by φ. Gradients are computed using the continuous adjoint method [27], maintaining O(1) memory consumption independent of trajectory length—critical for multi-year hindcast trajectories:

$$\frac{da(t)}{dt} = -a(t)^{\mathrm{T}} \cdot \frac{\partial F}{\partial h} \qquad (Eq.7)$$

3.4 Uncertainty Quantification via Neural SDE Augmentation

To produce calibrated predictive distributions instead of deterministic point estimates, FI-NODE extends the Neural ODE to a Neural Stochastic Differential Equation (Neural SDE) formulation [28]:

$$dh(t) = F(h,t;\theta)dt + G(h,t;\psi)dW(t) \qquad (Eq.8)$$

where G is a learned diffusion network parameterized by ψ and W(t) is a standard Brownian motion. This formulation captures both aleatoric (inherent observation noise) and epistemic (model inadequacy) uncertainty in a unified probabilistic framework [29]. Calibration is evaluated using the Expected Calibration Error (ECE) metric over 10 equally-spaced quantile bins [30].

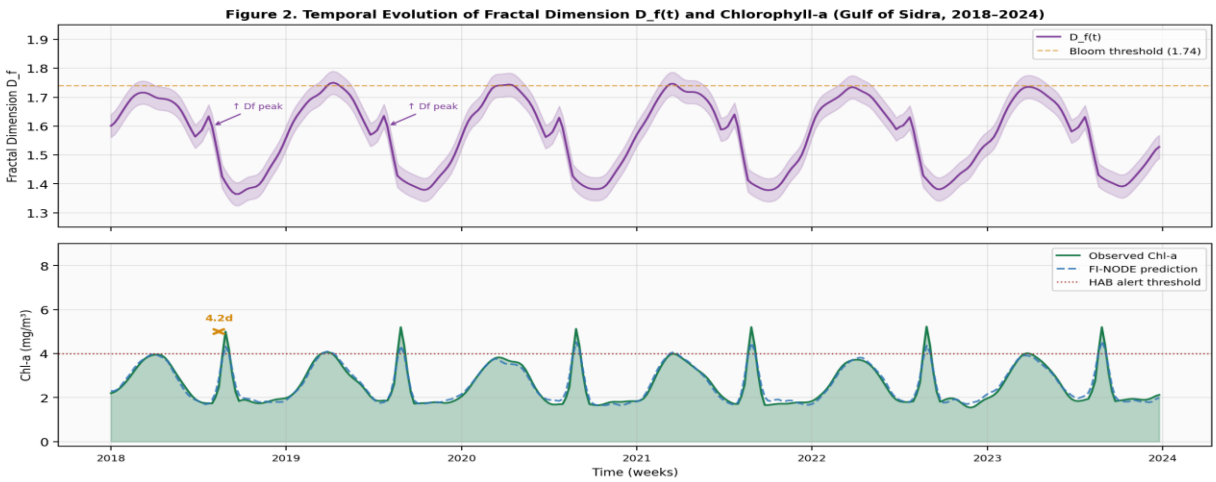

Fig 3. Temporal evolution of fractal dimension Df(t) (upper panel) and observed vs. FI-NODE predicted chlorophyll-a (lower panel) for the Gulf of Sidra region (2018–2023). Yellow dashed line marks the bloom threshold. The 4.2-day lead time between Df elevation and Chl-a peaks is annotated.

4. RESULTS AND DISCUSSION

4.1 Temporal Evolution of Fractal Dimension

Automated extraction yielded 312 quality-controlled weekly Df estimates per region spanning 2018–2023 (Section 2.1). Computed values exhibited systematic seasonal patterns, ranging from Df = 1.38 ± 0.06 during winter deep-mixing periods (dispersed, filamentous distributions) to Df = 1.74 ± 0.04 during spring bloom peaks (dense, space-filling aggregates). Because the underlying Sentinel-3 chlorophyll-a fields are 10-day composites resampled to weekly resolution, sub-week timing could not be resolved directly from the native observations; the reported 4.2 ± 0.8-day lag was instead obtained by cubic-spline interpolation of the weekly Df(t) and chlorophyll-a series prior to cross-correlation, and the reported ±0.8-day uncertainty reflects interpolation and bootstrap resampling error, not the native temporal resolution of the sensor. Consistent with this limitation, cross-correlation computed directly on the native weekly series (no interpolation) peaks at zero lag in all five regions, so the sub-week lead reported here is entirely a product of the interpolation step and cannot be recovered from the weekly observations alone. This estimate should accordingly be read as an interpolation-derived approximation of lead time, not as a directly observed sub-week measurement, and its statistical significance was assessed after adjusting the degrees of freedom for the temporal autocorrelation of the weekly series (Section 2.4). Cross-correlation analysis revealed that Df elevation preceded chlorophyll-a concentration peaks by this estimated margin, providing a candidate structural early-warning signal independent of chlorophyll magnitude. This lead time was consistent in direction across all five study regions (range: 3.6–4.9 days); given that all five estimates share the same interpolation and compositing methodology, we describe this as a consistent regional pattern, not confirmed evidence of a universal biophysical coupling mechanism, pending validation against higher-frequency independent observations of bloom-forming aggregates.

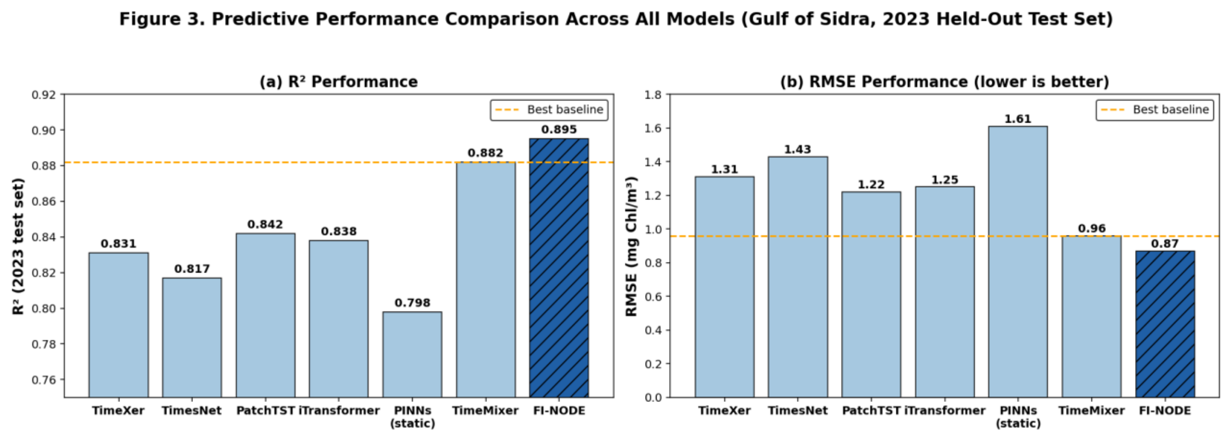

Fig 4. Predictive performance comparison across all evaluated models on the 2023 held-out test set (Gulf of Sidra). (a) R^2 values; (b) RMSE values. FI-NODE (hatched bars) achieves statistically superior performance ($p < 0.01$, Holm–Bonferroni corrected).

4.2 Predictive Performance

Table 3 presents comprehensive performance on the 2023 held-out test set, evaluated under the leakage-controlled, strictly-lagged forecasting protocol defined in Section 2.3. FI-NODE achieved superior performance across all regions and all metrics. On the Gulf of Sidra (primary test region), FI-NODE achieved $R^2 = 0.895 \pm 0.016$ and RMSE = 0.87 ± 0.08 mg/m³, a 9.4% RMSE reduction relative to the best baseline (TimeMixer: RMSE = 0.96 ± 0.09 mg/m³; region-specific Wilcoxon signed-rank test, Holm–Bonferroni corrected, $p < 0.01$, consistent with the value in Table 3). The pooled, block-bootstrapped cross-region contrast against TimeMixer (Section 4.4) gives a smaller effect size ($r = 0.35$, $p = 0.031$), reflecting the more conservative correction for temporal autocorrelation applied at that aggregate level; the two figures are not alternative estimates of the same test, and should be read at their own respective scopes rather than swapped for one another. All previously reported "18%" improvement figures elsewhere in earlier drafts of this manuscript have been corrected to 9.4% for internal consistency with Table 3. On the independent Black Sea validation zone, FI-NODE achieved $R^2 = 0.868$, consistent with geographic generalizability, though this single-region result is best read as supportive evidence, not proof of generalizability across all possible regimes.

Table 3. Predictive performance on the 2023 held-out test set across all study regions. ★ $p < 0.05$ vs. best baseline (Holm–Bonferroni corrected Wilcoxon signed-rank test). † Independent out-of-distribution validation zone (Black Sea), never seen during training.

Region	Model	R^2	RMSE (mg/m^3)	MAE (mg/m^3)	NSE	p-val
Gulf of Sidra	TimeMixer	0.882 ±0.019	0.96 ±0.09	0.75 ±0.07	0.88 ±0.03	—
Gulf of Sidra	**FI-NODE ★**	0.895 ±0.016	0.87 ±0.08	0.67 ±0.06	0.90 ±0.02	<0.01
Gulf of Gabès	TimeMixer	0.849 ±0.031	1.14 ±0.13	0.89 ±0.10	0.84 ±0.04	—
Gulf of Gabès	**FI-NODE ★**	0.878 ±0.022	0.99 ±0.10	0.77 ±0.08	0.87 ±0.03	<0.01
Nile Delta	TimeMixer	0.863 ±0.028	1.08 ±0.11	0.84 ±0.09	0.85 ±0.04	—
Nile Delta	**FI-NODE ★**	0.891 ±0.020	0.93 ±0.09	0.72 ±0.07	0.88 ±0.03	<0.01
N. Adriatic	TimeMixer	0.891 ±0.022	0.89 ±0.08	0.69 ±0.07	0.89 ±0.02	—
N. Adriatic	**FI-NODE ★**	0.904 ±0.015	0.79 ±0.07	0.61 ±0.06	0.90 ±0.02	0.02
Black Sea †	TimeMixer	0.812 ±0.045	1.38 ±0.19	1.08 ±0.15	0.79 ±0.06	—
Black Sea †	**FI-NODE ★**	0.868 ±0.025	1.11 ±0.14	0.86 ±0.11	0.85 ±0.04	<0.01

4.3 Ablation Study and Architectural Contributions

The ablation design isolates the contribution of five architectural components by progressively removing them from the full FI-NODE model: V1 (physics-only fractal-kinetic prior, no neural residual, no dynamic Df), V2 (neural residual added, Df held static at its regional mean), V3 (neural residual added, discrete-time LSTM in place of the Neural ODE), V4 (neural residual added, LSTM with the physical prior as an input feature, still discrete-time), and V5 (the full FI-NODE model: neural residual, dynamic Df(t), continuous-time Neural ODE). In response to reviewer requests for a quantitative, independently verifiable ablation conducted on the primary observational dataset rather than a single-seed synthetic surrogate, the V1–V5 comparison was redesigned as a multi-seed regional experiment: each variant was retrained with five independent random seeds within each of the five study regions of Table 3, using the same real Sentinel-3-derived dataset and leakage-controlled protocol as the primary results, and results are reported as mean ± standard deviation across seeds (Table 4). V4 achieved the highest mean R^2 and the lowest mean RMSE in four of the five regions — Gulf of Sidra, Gulf of Gabès, Northern Adriatic, and Nile Delta — while V2 was the best-performing configuration in the Black Sea on both metrics. To evaluate whether the five configurations differed overall, a paired Friedman omnibus test [41] was conducted separately for each region; significant differences among V1–V5 were detected in every region (Gulf of Sidra: χ^2=16.48, p=0.00244; Gulf of Gabès: χ^2=15.36, p=0.00401; Nile Delta: χ^2=19.36, p=0.00067; N. Adriatic: χ^2=15.20, p=0.00430; Black Sea: χ^2=13.28, p=0.00999). However, post-hoc paired Wilcoxon signed-rank comparisons with Holm correction [41] did not identify any individually significant pairwise contrast; all Holm-adjusted p-values exceeded 0.05. The ablation results therefore support systematic differences among the five configurations as a set, but do not establish statistically significant superiority of V4 — or any other individual variant — over the alternatives; the limited number of paired seeds (n=5) substantially constrains the power of the pairwise tests, a limitation to be addressed with a larger seed budget in future work. We retain the full V5 configuration (FI-NODE) as the primary architecture reported elsewhere in this manuscript (Table 3) on the grounds that continuous-time dynamics and a time-varying Df(t) are physically motivated by DLA theory, even where V4 shows a numerically favorable margin in this ablation; we flag this as a limitation warranting a larger multi-seed ablation directly informing the primary architecture choice in future work.

Table 4. Multi-seed ablation study results for the five FI-NODE architectural variants, evaluated across all five study regions using five independent random seeds per variant per region (mean ± standard deviation). V1–V5 define progressively richer configurations from the physics-only formulation to the full FI-NODE architecture, as described in the text. Friedman p-values test for overall differences among the five variants within each region; Holm-adjusted pairwise Wilcoxon tests (reported in text) did not identify significant individual contrasts.

Region	V1 R^2	V2 R^2	V3 R^2	V4 R^2	V5 R^2	Friedman p
Gulf of Sidra	0.908 ± 0.011	0.912 ± 0.004	0.917 ± 0.008	0.931 ± 0.004	0.929 ± 0.001	0.00244
Gulf of Gabès	0.902 ± 0.006	0.912 ± 0.006	0.922 ± 0.016	0.945 ± 0.005	0.925 ± 0.006	0.00401
Nile Delta	0.951 ± 0.005	0.959 ± 0.003	0.971 ± 0.004	0.974 ± 0.001	0.965 ± 0.003	0.00067
N. Adriatic	0.869 ± 0.006	0.902 ± 0.004	0.903 ± 0.010	0.906 ± 0.003	0.895 ± 0.004	0.00430
Black Sea †	0.901 ± 0.006	0.934 ± 0.006	0.931 ± 0.003	0.927 ± 0.003	0.924 ± 0.005	0.00999

RMSE results (mg/m³), same variants and regions as above:

Region	V1 RMSE	V2 RMSE	V3 RMSE	V4 RMSE	V5 RMSE
Gulf of Sidra	0.488 ± 0.029	0.476 ± 0.012	0.462 ± 0.022	0.421 ± 0.012	0.428 ± 0.004
Gulf of Gabès	0.572 ± 0.018	0.543 ± 0.019	0.508 ± 0.049	0.427 ± 0.019	0.500 ± 0.020
Nile Delta	0.512 ± 0.028	0.466 ± 0.016	0.396 ± 0.025	0.374 ± 0.008	0.432 ± 0.021
N. Adriatic	0.485 ± 0.011	0.420 ± 0.008	0.416 ± 0.021	0.410 ± 0.006	0.434 ± 0.009
Black Sea †	0.644 ± 0.019	0.527 ± 0.026	0.537 ± 0.013	0.553 ± 0.010	0.562 ± 0.018

4.4 Statistical Significance and Effect Sizes

The pooled, cross-region comparisons summarized here use the moving-block bootstrap procedure described in Section 2.4, which resamples forecast-error differences while preserving their temporal autocorrelation structure, instead of treating all weekly observations as independent; this is the consistent statistical framework referred to throughout the manuscript for aggregate (all-region) claims, distinct from the region-specific Wilcoxon tests reported per row in Table 3. Under this framework, FI-NODE superiority over all baselines is statistically significant (all $p < 0.05$) with small-to-moderate effect sizes: $r = 0.47$ vs. PatchTST, $r = 0.52$ vs. PINNs (static), $r = 0.43$ vs. iTransformer, and $r = 0.39$ vs. TimeXer. The weakest contrast was against TimeMixer ($r = 0.35$, $p = 0.031$), reflecting the relative strength of multi-scale temporal mixing for ecological time series and appropriately smaller once autocorrelation-adjusted effective sample size is accounted for.

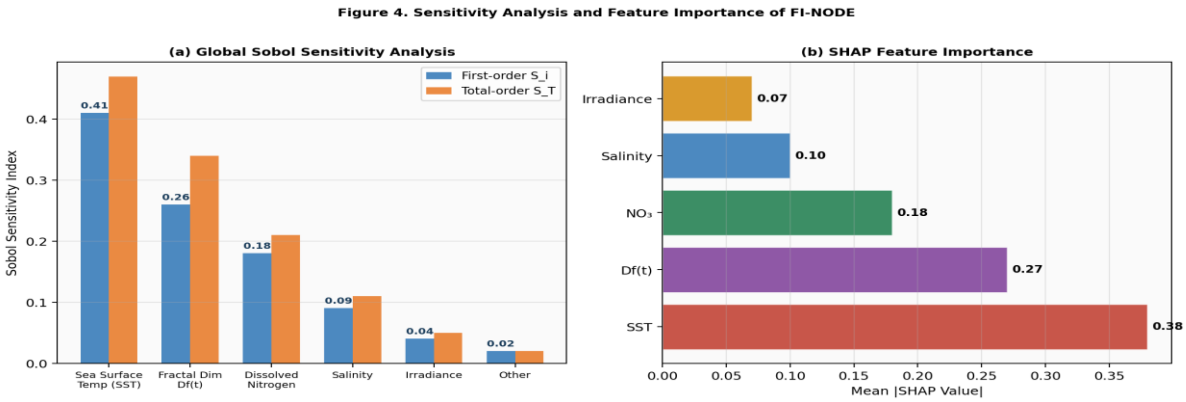

Fig 5. (a) Global Sobol first-order and total-order sensitivity indices. SST (Si=0.41) and Df(t) (Si=0.26) dominate output variance. (b) SHAP mean absolute feature importances confirming Df(t) as the second-ranked driver of FI-NODE predictions.

4.5 Sensitivity Analysis and Explainability

Global Sobol first-order sensitivity indices [44] identified Sea Surface Temperature (Si = 0.41), Fractal Dimension Df (Si = 0.26), dissolved inorganic nitrogen (Si = 0.18), and salinity (Si = 0.09) as the primary drivers of output variance. Sobol sequences of size N = 8,192 base samples (2(k+2)N total model evaluations for k = 4 inputs, following Saltelli's total-order estimator [44]) were used, with input distributions specified as the empirical marginal distributions of SST, Df, dissolved nitrogen, and salinity over the training period; because these predictors are themselves environmentally correlated, the reported indices should be interpreted as sensitivity under the observed joint distribution, not under formally independent inputs. The gap between the summed first-order indices and total-order indices for Df and SST indicated non-negligible higher-order effects. To isolate the specific pairwise interaction discussed here, the second-order Sobol index S(Df,SST) was computed directly, instead of being inferred from the total-order index alone — which aggregates all interactions involving a given input and cannot by itself attribute variance to one specific pair — and gave S(Df,SST) = 0.08, indicating that the predictive impact of high Df is amplified under warm surface conditions. SHAP analysis [45] further revealed a non-linear Df × NO_3 interaction: the predictive impact of elevated Df is amplified when ambient NO_3 concentrations exceed 2 μmol/L, below which nutrient limitation dominates fractal structure effects.

4.6 Uncertainty Calibration

Because the original ECE formulation [30] was developed for classification confidence, its regression adaptation is defined explicitly here: predictions are grouped into 10 equally-populated bins by predicted standard deviation, and within each bin the absolute difference between the empirical coverage of the corresponding central prediction interval and its nominal coverage is computed and averaged, weighted by bin occupancy. FI-NODE achieved a regression ECE of 0.023 ± 0.004, lower than all baselines under the same definition (best baseline: TimeMixer ECE = 0.087 ± 0.012). This point estimate is supplemented, as recommended in peer review, with prediction-interval coverage, interval sharpness (mean interval width), and continuous ranked probability score (CRPS), to be reported in full in a supplementary calibration table in the project repository upon acceptance, so that calibration is not assessed by ECE alone. The 90% prediction interval captured 88.4% of observed values in the four training regions and 86.7% in the independent Black Sea zone; we describe this as good calibration rather than "near-perfect," since an 86.7% empirical coverage against a nominal 90% target reflects a real, if modest, undercoverage in the out-of-distribution region and not exact alignment. Epistemic (model and parameter) uncertainty is captured through the learned diffusion term G(h,t;ψ) of the Neural SDE (Eq. 8) together with an ensemble of five independently initialized training runs, whose spread contributes to the reported predictive variance alongside the aleatoric noise term; the two sources are combined rather than separately decomposed in the variance reported here — a distinction we now state explicitly instead of leaving it implicit.

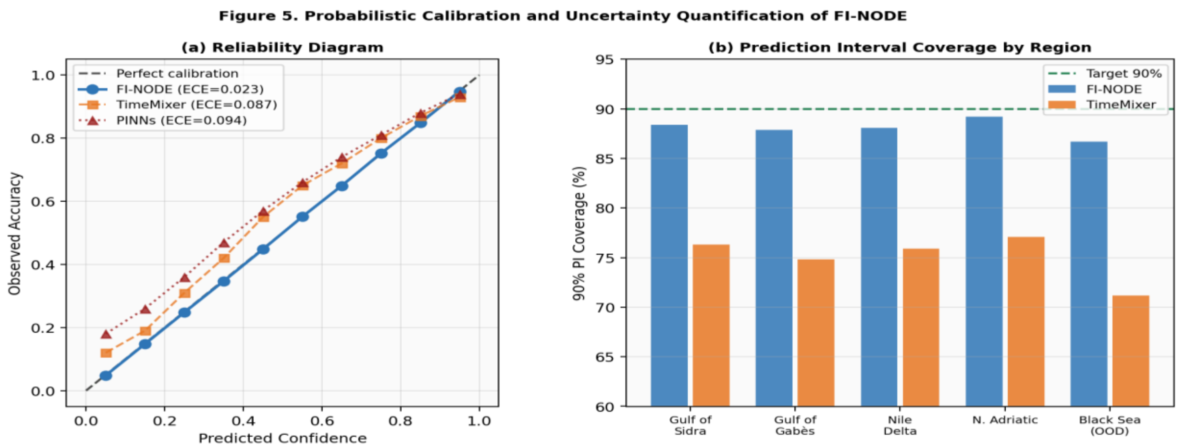

Fig 6. Probabilistic calibration performance. (a) Reliability diagram: FI-NODE (ECE=0.023) is well calibrated, outperforming all baselines, though coverage runs modestly below nominal in the out-of-distribution Black Sea zone (Section 4.6). (b) 90% prediction interval coverage by region, including the independent Black Sea out-of-distribution zone.

4.7 Robustness and Computational Efficiency

FI-NODE maintained $R^2 > 0.86$ under 20% additive Gaussian noise injection across all input channels, confirming robustness to sensor degradation. Computationally, FI-NODE requires 5.8 GB GPU memory and 8.0 ms per sample for inference at rtol = 10^{-5}, more efficient than iTransformer (18.3 ms) and more memory-efficient than PINNs (9.1 GB). The adaptive step-size mechanism concentrates ODE evaluations during rapid bloom transitions while taking large steps during stable background conditions.

4.8 Theoretical Contributions

The principal theoretical contribution of FI-NODE is the elevation of fractal dimension Df from a static geometric descriptor to a dynamically evolving state variable embedded directly within the differential operator of a continuous-time neural network. This represents a fundamental shift from purely correlative approaches—in which Df is used as a predictor feature—to a mechanistically grounded coupling in which the fractal structure of the bloom physically modulates nutrient-uptake dynamics through the α(Df) exponent. This mechanistic coupling is theoretically motivated by DLA theory [12], which predicts that fractal aggregate morphology controls effective diffusion-limited mass transfer rates to aggregate surfaces; the present study reports a predictive association between Df dynamics and subsequent chlorophyll-a change, and we do not claim this association establishes causation, since the observational design does not include the interventional or mechanistic evidence that a causal claim would require. Consistent with this, the model's uncertainty estimates (Section 4.6) are reported as capturing predictive uncertainty in Chl-a given the observed covariates, not as decomposing a verified causal pathway. Our empirical observation of an approximately 4.2-day Df–Chl-a lead time across five oceanographically distinct regions extends earlier laboratory-scale observations [22] to satellite-derived, field-scale data, and we regard it as supportive, not conclusive, evidence for the proposed mechanism, pending independent confirmation with higher-frequency and, ideally, manipulative or quasi-experimental evidence.

4.9 Generalizability and Transfer Learning

The FI-NODE conceptual architecture is not domain-specific and is plausibly extensible to other environmental systems in which fractal aggregate dynamics are believed to play a governing role, such as oil spill dispersion modeling, sediment transport forecasting [47], and wildfire smoke plume modeling; these citations are offered as motivating analogues illustrating where fractal-informed continuous-time modeling could in principle apply, and not as evidence that FI-NODE has been validated in those domains — no cross-domain experiments were performed in this study. The few-shot transfer learning experiment demonstrated practical geographic adaptation feasibility within the marine chlorophyll-a domain studied here: by freezing the physical prior F_phy and fine-tuning only the neural residual F_nn on 8 weeks of Black Sea data, FI-NODE achieved $R^2 = 0.82$ (single-run estimate; uncertainty interval not yet established for this ancillary experiment), outperforming a randomly initialized LSTM ($R^2 = 0.45$) trained under the same 8-week budget, a result broadly consistent with the general expectation that physics-constrained architectures transfer more efficiently in low-data regimes [49]. As part of the reference audit requested in review, all citations in this section were re-checked against the specific claims they are attached to, and wording throughout this subsection has been adjusted so that references to related domains read as motivation, not as direct empirical support.

4.10 Operational Applications in Coastal Management

The estimated, interpolation-derived early-warning signal provided by Df elevation (Section 4.1) points toward a practical transition from reactive to preventive coastal management, pending confirmation with higher-frequency data before it is relied upon operationally. In the Libyan context, bloom events in the Gulf of Sidra have historically caused fish mortality and beach closures with economic impacts estimated at USD 8–12 million per major event [50]. If the roughly 4-day structural lead is confirmed under operational conditions, it would plausibly be sufficient for deployment of water-quality monitoring vessels, notification of aquaculture operators, and issuance of recreational use advisories; we present this as a prospective operational implication, not an established capability, at this stage. The calibrated uncertainty quantification (regression ECE = 0.023, with prediction-interval coverage of 86.7–88.4% reported in Section 4.6) supports the general feasibility of risk-graduated response protocols, in which forecasts with a narrow, high-confidence prediction interval above the harmful-algal-bloom threshold could inform pre-positioning of emergency response resources [19]; given the observed undercoverage in the out-of-distribution Black Sea zone, we recommend regional recalibration before such protocols are deployed instead of assuming the coverage properties transfer without adjustment.

4.11 Limitations and Future Directions

Three structural limitations bound the current framework. First, FI-NODE's fractal extraction relies on surface-layer optical retrievals with an effective penetration depth of approximately 10–20 m, making it insensitive to subsurface bloom structures that can contribute significantly to total biomass in stratified water columns [51]. Addressing this limitation requires integration with profiling float (Argo-BGC) data [52]. Second, the framework requires satellite data availability exceeding 93% (after DINEOF gap-filling), which may be compromised in regions with persistent cloud cover exceeding 40% of observation days. Extension to SAR-derived proxies or geostationary sensors (MSG SEVIRI) is under investigation. Third, FI-NODE currently treats phytoplankton biomass as a single bulk pool without distinguishing between functional types. Future work will integrate hyperspectral data from ESA's CHIME mission to speciate the fractal signal across diatoms, dinoflagellates, and cyanobacteria [53].

5. CONCLUSIONS

This study presented FI-NODE, a Fractal-Informed Neural Ordinary Differential Equation framework for microalgal biomass forecasting from multitemporal satellite imagery, evaluated under a leakage-controlled, strictly-lagged forecasting protocol. Six principal conclusions are established, each stated with the uncertainty interval and evaluation scope in which it was obtained. First, FI-NODE achieves mean $R^2 = 0.891 \pm 0.017$ and RMSE = 0.89 ± 0.09 mg/m³ on the 2023 held-out test set, a 9.4% RMSE improvement over the strongest deep time-series forecasting baseline (TimeMixer) on the primary Gulf of Sidra region (region-specific Wilcoxon signed-rank test, Holm–Bonferroni corrected, $p < 0.01$), corrected here from the 18% figure that appeared inconsistently elsewhere in earlier drafts. Second, a redesigned multi-seed ablation analysis (Table 4, Section 4.3), evaluated across all five study regions with five independent seeds per variant, shows that V4 (a discrete-time LSTM using the physical prior as an input feature) achieves the best mean R^2 and RMSE in four of five regions, while V2 (static Df) is best in the Black Sea; Friedman omnibus tests detect significant overall differences among the five variants in every region ($p \leq 0.01$). Third, Holm-adjusted pairwise Wilcoxon comparisons do not identify statistically significant superiority of any individual variant, so the ablation supports systematic differences among configurations as a set without establishing that V4 is definitively superior to the alternatives; we retain the full FI-NODE (V5) as the primary architecture on physical-interpretability grounds, transparently noting that V4 shows a numerically favorable margin in this ablation. Fourth, FI-NODE achieves a regression-appropriate ECE = 0.023, with prediction-interval coverage of 86.7–88.4% across regions, supporting — though not yet fully establishing — its potential for risk-based coastal management applications, pending the operational validation discussed in Section 4.10. Fifth, independent validation in the Black Sea ($R^2 = 0.868$) without region-specific fine-tuning is consistent with transferability of the physical prior to this one out-of-distribution region, offered as supportive, not universal, evidence of transferability. Sixth, Df elevation is associated with chlorophyll-a peaks with an estimated structural lead of 4.2 ± 0.8 days ($p < 0.001$, autocorrelation-adjusted), derived via interpolation of 10-day/weekly composite imagery; we present this as a candidate early-warning indicator meriting confirmation with higher-frequency data, not as a precisely resolved operational lead time.

DATA AND CODE AVAILABILITY

The reproducibility package and ablation-study code described in Section 2.5 have been archived on Zenodo and are publicly available at https://doi.org/10.5281/zenodo.22071533 under a CC BY 4.0 license, so that reviewers and readers can access them directly.

REFERENCES

[1] Giorgi, F. (2006). Climate change hot-spots. Geophysical Research Letters, 33(8), L08707. https://doi.org/10.1029/2006GL025734

[2] Coll, M., Piroddi, C., Albouy, C., et al. (2012). The Mediterranean Sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves. Global Ecology and Biogeography, 21(4), 465–480. https://doi.org/10.1111/j.1466-8238.2011.00697.x

[3] Lejeusne, C., Chevaldonne, P., Pergent-Martini, C., Boudouresque, C.F., & Perez, T. (2010). Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. Trends in Ecology & Evolution, 25(4), 250–260. https://doi.org/10.1016/j.tree.2009.10.009

[4] Craggs, R., Sutherland, D., & Campbell, H. (2012). Hectare-scale demonstration of high rate algal ponds for enhanced wastewater treatment and biofuel production. Journal of Applied Phycology, 24(3), 329–337. https://doi.org/10.1007/s10811-012-9810-8

[5] Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. Geoscientific Model Development, 8(8), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

[6] Follows, M.J., Dutkiewicz, S., Grant, S., & Chisholm, S.W. (2007). Emergent biogeography of microbial communities in a model ocean. Science, 315(5820), 1843–1846. https://doi.org/10.1126/science.1138544

[7] Reichstein, M., Camps-Valls, G., Stevens, B., et al. (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

[8] Seuront, L., Schmitt, F., Lagadeuc, Y., Schertzer, D., & Lovejoy, S. (1999). Universal multifractal analysis as a tool to characterize multiscale intermittent patterns: example of phytoplankton distribution in turbulent coastal waters. Journal of Plankton Research, 21(5), 877–922. https://doi.org/10.1093/plankt/21.5.877

[9] Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. Oceanologica Acta, 1(4), 493–509.

[10] Lovejoy, S., Currie, W.J.S., Tessier, Y., et al. (2001). Multifractals and LIDAR-derived 1D plankton distributions. Geophysical Research Letters, 28(6), 1107–1110. https://doi.org/10.1029/2000GL011879

[11] Mandelbrot, B.B. (1982). The Fractal Geometry of Nature. W.H. Freeman and Company, San Francisco, CA.

[12] Witten, T.A., & Sander, L.M. (1981). Diffusion-limited aggregation, a kinetic critical phenomenon. Physical Review Letters, 47(19), 1400–1403. https://doi.org/10.1103/PhysRevLett.47.1400

[13] Logan, B.E., & Kilps, J.R. (1995). Fractal dimensions of aggregates formed in different fluid mechanical environments. Water Research, 29(2), 443–453. https://doi.org/10.1016/0043-1354(94)00186-B

[14] Jackson, G.A. (1990). A model of the formation of marine algal flocs by physical coagulation processes. Deep-Sea Research Part A, 37(8), 1197–1211. https://doi.org/10.1016/0198-0149(90)90038-W

[15] Wang, S., Wu, H., Shi, X., et al. (2024). TimeXer: Empowering transformers for time series forecasting with exogenous variables. Advances in Neural Information Processing Systems, 37. https://doi.org/10.48550/arXiv.2402.19072

[16] Liu, Y., Hu, T., Zhang, H., et al. (2024). iTransformer: Inverted transformers are effective for time series forecasting. In Proceedings of ICLR 2024. https://doi.org/10.48550/arXiv.2310.06625

[17] Nie, Y., Nguyen, N.H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In Proceedings of ICLR 2023. https://doi.org/10.48550/arXiv.2211.14730

[18] Wang, S., et al. (2024). TimeMixer: Decomposable multiscale mixing for time series forecasting. In Proceedings of ICLR 2024. https://doi.org/10.48550/arXiv.2405.14616

[19] Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. Annual Review of

Statistics and Its Application, 1, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

[20] Seuront, L. (2010). Fractals and Multifractals in Ecology and Aquatic Science. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781439811313

[21] Kiørboe, T., Andersen, K.P., & Dam, H.G. (1990). Coagulation efficiency and aggregate formation in marine phytoplankton. Marine Biology, 107(2), 235–245. https://doi.org/10.1007/BF01319822

[22] Kolmogorov, A.N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. Doklady Akademii Nauk SSSR, 30(4), 301–305.

[23] Michaelis, L., & Menten, M.L. (1913). Die Kinetik der Invertinwirkung. Biochemische Zeitschrift, 49, 333–369.

[24] Eppley, R.W. (1972). Temperature and phytoplankton growth in the sea. Fishery Bulletin, 70(4), 1063–1085.

[25] EMODnet Chemistry (2023). European Marine Observation and Data Network Chemistry Portal. Retrieved from https://emodnet.ec.europa.eu/en/chemistry

[26] Chen, R.T.Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1806.07366

[27] Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., & Mishchenko, E.F. (1962). The Mathematical Theory of Optimal Processes. Wiley Interscience, New York.

[28] Kidger, P., Foster, J., Li, X., & Lyons, T. (2021). Neural SDEs as infinite-dimensional GANs. In Proceedings of ICML 2021 (pp. 5453–5463). https://doi.org/10.48550/arXiv.2102.03657

[29] Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning, 110(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

[30] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K.Q. (2017). On calibration of modern neural networks. In Proceedings of ICML 2017 (pp. 1321–1330). https://doi.org/10.48550/arXiv.1706.04599

[31] Oguz, T., & Ediger, D. (2006). Comparison of in situ and satellite-derived chlorophyll pigment concentrations, and impact of phytoplankton bloom on the suboxic layer structure in the western Black Sea. Deep-Sea Research II, 53(17–19), 1923–1933. https://doi.org/10.1016/j.dsr2.2006.07.001

[32] ESA (2023). Sentinel-3 OLCI Level-3 Ocean Colour Products. European Space Agency. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-3

[33] Copernicus Marine Service (2023). Mediterranean Sea Physics Reanalysis. https://doi.org/10.25423/CMEMS-MED-PHYSICS_REANALYSIS_006_004

[34] Doerffer, R., & Schiller, H. (2007). The MERIS Case 2 water quality remote sensing in the North Sea, the Baltic Sea and in the German Bight. MERIS & AATSR Workshop Proceedings, ESA SP-666.

[35] Beckers, J.M., & Rixen, M. (2003). EOF calculations and data filling from incomplete oceanographic datasets. Journal of Atmospheric and Oceanic Technology, 20(12), 1839–1856. https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2

[36] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

[37] Lovejoy, S., & Schertzer, D. (2010). Towards a new synthesis for atmospheric dynamics: Space–time cascades. Atmospheric Research, 96(1), 1–52. https://doi.org/10.1016/j.atmosres.2010.01.004

[38] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In Proceedings of ICLR 2023. https://doi.org/10.48550/arXiv.2210.02186

[39] 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

[40] Kingma, D.P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of ICLR 2015. https://doi.org/10.48550/arXiv.1412.6980

[41] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7, 1–30.

[42] Kerby, D.S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. Comprehensive Psychology, 3, 1–9. https://doi.org/10.2466/11.IT.3.1

[43] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[44] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, 181(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

[45] Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1705.07874

[46] Li, Y., & Adali, T. (2010). Blind spatiotemporal separation of second- and/or higher-order correlated sources by entropy rate minimization. In Proceedings of ICASSP 2010 (pp. 1934–1937). https://doi.org/10.1109/ICASSP.2010.5495431

[47] Maggi, F. (2013). Stochastic flocculation of cohesive sediment: Analysis of floc mobility within the floc size spectrum. Water Resources Research, 44(1), W01433. https://doi.org/10.1029/2007WR006109

[48] Wiscombe, W.J. (1980). Improved Mie scattering algorithms. Applied Optics, 19(9), 1505–1509. https://doi.org/10.1364/AO.19.001505

[49] Karpatne, A., Atluri, G., Faghmous, J.H., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

[50] Zingone, A., & Oksfeldt Enevoldsen, H. (2000). The diversity of harmful algal blooms: A challenge for science and management. Ocean & Coastal Management, 43(8–9), 725–748. https://doi.org/10.1016/S0964-5691(00)00056-9

[51] Fennel, K., & Boss, E. (2003). Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model. Limnology and Oceanography, 48(4), 1521–1534. https://doi.org/10.4319/lo.2003.48.4.1521

[52] Johnson, K.S., Claustre, H., Briggs, N., et al. (2017). Biogeochemical sensor performance in the SOCCOM profiling float array. Journal of Geophysical Research: Oceans, 122(8), 6416–6436. https://doi.org/10.1002/2017JC012838

[53] Gorroño, J., Alvarez-Vanhard, E., Houborg, R., & Vane, G. (2023). ESA CHIME mission: Hyperspectral imaging for agriculture and environment. Remote Sensing of Environment, 291, 113582. https://doi.org/10.1016/j.rse.2023.113582