

Article

Not peer-reviewed version

Physical and Biogeochemical Drivers for Forecasting Red Tides in Southwest Florida: A Regionally Integrated Machine Learning Framework

Matthew Duus , [Ahmed S. Elshall](#) ^{*} , Michael L. Parsons , [Ming Ye](#)

Posted Date: 6 March 2026

doi: 10.20944/preprints202603.0568.v1

Keywords: bloom forecasting; *Karenia brevis*; red tide; harmful algal blooms; nutrient loading; watershed-coastal coupling; Gulf of Mexico or Gulf of America; machine learning; random forest classifier



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Physical and Biogeochemical Drivers for Forecasting Red Tides in Southwest Florida: A Regionally Integrated Machine Learning Framework

Matthew Duus ¹, Ahmed S. Elshall ^{1,2,3,*}, Michael L. Parsons ² and Ming Ye ⁴

¹ Department of Bioengineering, Civil Engineering, and Environmental Engineering, U.A. Whitaker College of Engineering, Florida Gulf Coast University, Fort Myers, Florida

² The Water School, Florida Gulf Coast University, Fort Myers, Florida

³ DENDRITIC: A Human-Centered Artificial Intelligence and Data Science Institute, Florida Gulf Coast University, Fort Myers, Florida

⁴ Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee Florida

* Correspondence: aelshall@fgcu.edu

Key points

- Integrates multi-watershed data with shelf-scale physical forcing
- Weekly red tide forecasting achieved ~90% accuracy with robust bloom detection skill
- Improves regional model spatial generalization by including Tampa Bay records

Abstract

Harmful algal blooms (HABs) caused by *Karenia brevis* (*K. brevis*) present a persistent ecological and public health challenge across coastal Florida. This study develops a regionally integrated machine learning framework to predict weekly *K. brevis* bloom occurrence using environmental data from both the Peace and Caloosahatchee Rivers, combined with coastal bloom records from Southwest Florida and Tampa Bay to enhance the spatial and temporal continuity of the response record. A Random Forest classifier was trained on a multi-decadal dataset incorporating river discharge, nutrient concentrations (total nitrogen and total phosphorus), wind forcing, sea surface temperature, salinity, and sea surface height anomalies as a proxy for Loop Current variability. The model achieved strong predictive performance on a chronologically withheld test set, with an overall accuracy of ~90%, balanced accuracy of 87.6%, and high precision and recall for bloom events. Bloom timing and persistence were captured with strong agreement during ongoing bloom periods, while non-bloom conditions were identified with low false-positive rates. Feature-response analyses indicated that bloom probability increased most sharply under moderate discharge and nutrient conditions, with diminished sensitivity at higher extremes. Learning curve analysis demonstrated robust training performance and stable generalization, with validation accuracy plateauing near 84%, suggesting a data-limited ceiling on forecast skill. By aggregating nutrient inputs across multiple watersheds and integrating spatially aligned bloom observations, this study demonstrates the utility of multi-source machine learning frameworks for regional-scale HAB prediction. The results support the development of early warning tools and provide a reproducible foundation for evaluating how combined watershed loading and physical forcing are associated with *K. brevis* bloom occurrence in complex estuary systems with watershed and coastal coupling.

Keywords: bloom forecasting; *Karenia brevis*; red tide; harmful algal blooms; nutrient loading; watershed–coastal coupling; Gulf of Mexico or Gulf of America; machine learning; random forest classifier

1. Introduction

Coastal water quality degradation driven by nutrient enrichment and harmful algal blooms (HABs) is an increasingly urgent environmental and public health concern worldwide (Griffith & Gobler, 2020; Wells et al., 2015). Along the West Florida Shelf, blooms of *Karenia brevis* (*K. brevis*), commonly referred to as red tide, are among the most persistent and damaging HABs, producing brevetoxins that cause respiratory illness, widespread fish kills, and substantial economic losses to coastal communities (Marcillo-Yeppez et al., 2025; Medina et al., 2022; Zheng et al., 2025; Zohdi & Abbaspour, 2019). Although *K. brevis* blooms are a natural feature of the Gulf of Mexico, mounting evidence indicates that anthropogenic nutrient enrichment and altered hydrologic regimes may intensify bloom persistence and coastal impacts (Elshall et al., 2020; Glibert et al., 2026; Medina et al., 2022; Vargo et al., 2008). Two major river systems, the Peace River and the Caloosahatchee River, exert strong influence on coastal water quality in southwest Florida by delivering freshwater, nutrients, and organic matter to nearshore waters. These rivers drain watersheds characterized by intensive agriculture, urban development, and managed flow control structures, creating highly variable discharge and nutrient loading conditions (Yan et al., 2024). Previous studies have documented associations between riverine nutrient inputs, particularly total nitrogen (TN) and total phosphorus (TP), and *K. brevis* bloom development, though the relative importance of individual watersheds and nutrients remains debated (Lenes & Heil, 2010; Vargo et al., 2008). Importantly, nutrient delivery does not act in isolation but interacts with physical drivers such as wind forcing, shelf circulation, and Loop Current variability to regulate bloom transport, retention, and persistence (Heil, Bronk, et al., 2014; Heil, Dixon, et al., 2014).

Recent advances in machine learning (ML) have provided practical tools for modeling HABs in complex coastal systems, where nonlinear interactions and class imbalance limit the effectiveness of traditional statistical approaches (Elshall et al., 2022). ML classifiers trained on physical and hydrologic drivers including Loop Current cycles, wind forcing, and river discharge have demonstrated a skill in distinguishing between large bloom and non-bloom conditions along the West Florida Shelf (Elshall et al., 2021). A similar framework (Li et al., 2021) has integrated cumulative nutrient loading. While previous studies (Elshall et al., 2021; Li et al., 2021) have prioritized environmental drivers, recent research (Medina et al., 2024; Yan et al., 2024) suggests that incorporating autoregressive features (lagged bloom observations) can increase operational accuracy by capturing bloom persistence, though often at the expense of isolating the contribution of external environmental forcing.

However, distinct challenges remain regarding model sensitivity and spatial transferability. First, reliance on recent bloom history can bias predictions toward persistence, potentially obscuring the environmental drivers responsible for true bloom onset following non-bloom periods (Medina et al., 2024; Yan et al., 2024). Second, models trained on restricted geographic domains may struggle to generalize to new regions. Specifically, it remains to be determined whether expanding the training domain of a localized Southwest Florida model by explicitly incorporating bloom records from adjacent coastal regimes such as Tampa Bay enhances predictive robustness and spatial generalization (Medina et al., 2024). To address these needs, this study develops a regionally integrated machine learning framework to predict *K. brevis* bloom occurrence along the Southwest Florida coast. Building on existing ML approaches (Elshall et al., 2021; Li et al., 2021), the framework integrates watershed-scale nutrient concentrations (TN and TP) and discharge from both the Peace and Caloosahatchee Rivers, alongside atmospheric forcing, Loop Current variability, and local hydrographic conditions. Bloom observations from the Tampa Bay region are incorporated to expose the model to a broader range of bloom and non-bloom transitions and to evaluate whether such expansion improves predictive robustness for the Southwest Florida domain. Lagged bloom predictors are included to represent short-term biological persistence rather than to resolve bloom initiation mechanisms (Yan et al., 2024), while the relative predictive influence of cumulative nutrient concentrations and physical forcing is quantified under these conditions. By synthesizing these data

sources, this study provides a reproducible framework for evaluating environmental conditions associated with regional *K. brevis* bloom occurrence.

2. Methods

2.1. Study Area and Bloom Observations

The study domain (Figure 1) focuses on the southwest Florida coastal region, specifically nearshore waters influenced by the Peace River and Caloosahatchee River watersheds, extending from Charlotte Harbor to the Tampa Bay region (Heil, Dixon, et al., 2014). This area was selected to capture coastal environments most directly impacted by terrestrial nutrient loading and shelf-scale physical forcing (Brand et al., 2012; Brand & Compton, 2007). Observations of *K. brevis* cell concentrations were obtained from the Florida Fish and Wildlife Conservation Commission (FWC-FWRI) long-term monitoring database for the period between January 1990 and December 2024 (FWC-FWRI, 2025a). To align *K. brevis* observations with environmental predictors, cell counts were aggregated to a weekly temporal resolution. For the machine-learning classification task, bloom presence was defined using a binary threshold of $\geq 100,000$ cells L^{-1} consistent with operational monitoring criteria and prior ecological studies (Brand et al., 2012; FWC-FWRI, 2025b; Heil, Dixon, et al., 2014). Observations below this threshold were classified as non-bloom conditions.

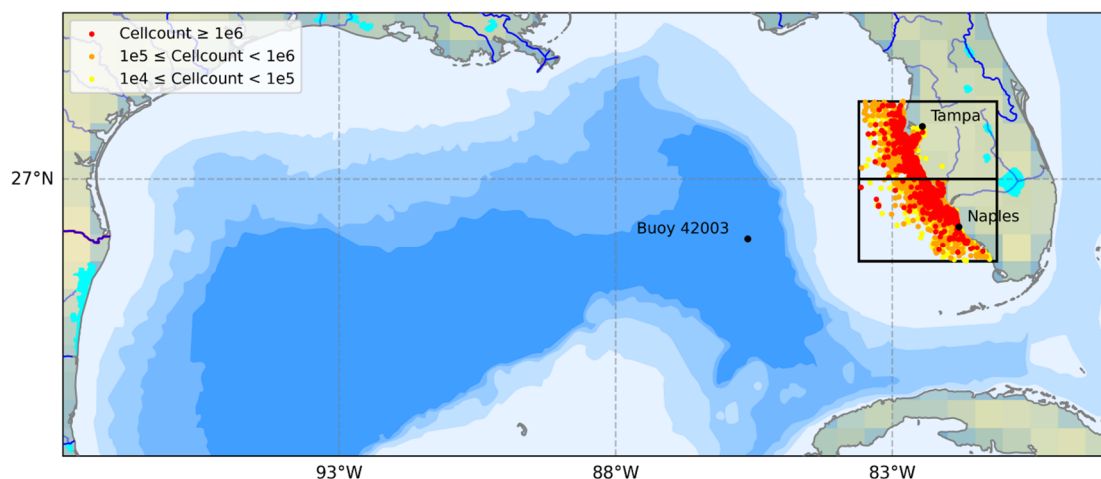


Figure 1. Regional study domain and coastal context. Map of the southwest Florida coastline illustrating the spatial distribution of *K. brevis* monitoring locations relative to coastal bathymetry and the primary discharge points of the Peace River and Caloosahatchee River watersheds.

2.2. Physical Drivers

Sea surface temperature (SST) and salinity data were obtained from the same FWC-FWRI monitoring records used for the biological data, ensuring temporally aligned sampling (FWC-FWRI, 2025a). These variables were aggregated to weekly maximum values to preserve episodic extremes, such as seasonal warming or freshwater pulses, that influence the physical and biogeochemical environment (Brand & Compton, 2007; Chen & Hu, 2017). Wind speed and direction data were sourced from NOAA buoy 42003 on the West Florida Shelf (NDBC, 2025). Large-scale circulation variability associated with the Loop Current was represented using sea surface height (SSH) anomalies (zos). Daily gridded SSH data were obtained from the Copernicus Marine Environment Monitoring Service (CMEMS) global ocean physics product (GLOBAL_MULTIYEAR_PHY_001_030) at an ~ 8 km horizontal resolution (Dréville et al., 2018; Fernandez & Lellouche, 2018). SSH anomalies were spatially averaged across the Gulf of Mexico domain and detrended to isolate high-frequency fluctuations associated with Loop Current excursions and eddy activity (Weisberg et al.,

2014). Both contemporaneous and lagged SSH values were included to account for the delayed physical response of coastal waters to basin-scale circulation forcing.

Wind vectors were converted to zonal (u) and meridional (v) components and vector-averaged to weekly mean values. Lagged wind predictors were then constructed by shifting these weekly time series by one- and two-week intervals to account for the delayed response of surface currents and bloom transport to atmospheric forcing, lagged wind variables were computed at weekly intervals, reflecting documented wind-driven circulation timescales (Basterretxea et al., 2024; Maze et al., 2015).

2.3. River Discharge and Nutrient Loading

Land-based freshwater and nutrient inputs were quantified for the Peace and Caloosahatchee Rivers, the primary sources of nitrogen and phosphorus to the region (Heil, Bronk, et al., 2014; Medina et al., 2022). For the Peace River (Figure 2a), daily discharge data were obtained from U.S. Geological Survey (USGS) gauging stations near Bartow and Arcadia, and corresponding total nitrogen (TN) and total phosphorus (TP) concentrations were retrieved from the Water Atlas of Florida (USF Water Institute, 2026). For the Caloosahatchee River (Figure 2b), discharge data were sourced from the S-79 flow control structure, which serves as the primary freshwater delivery point to the estuary. Nutrient data (TN and TP) for the Caloosahatchee were obtained from the Florida Storage and Retrieval (STORET) database and the Water Quality Portal (Florida Department of Environmental Protection, 2025). Details about data curation procedures for discharge and nutrient data of Peace River and Caloosahatchee River are included in the supplementary material (Duus, 2026; A. S. Elshall, 2025). For both river systems, discharge and nutrient concentration data were resampled to a weekly resolution. Weekly maximum values were retained to capture the episodic high-flow pulses and high-concentration that have been shown to exert a disproportionate influence on coastal phytoplankton dynamics (Heil, Dixon, et al., 2014; Medina et al., 2022; Glibert et al., 2026). Lagged predictors for discharge and nutrients were also constructed to represent delayed biological responses to watershed forcing.

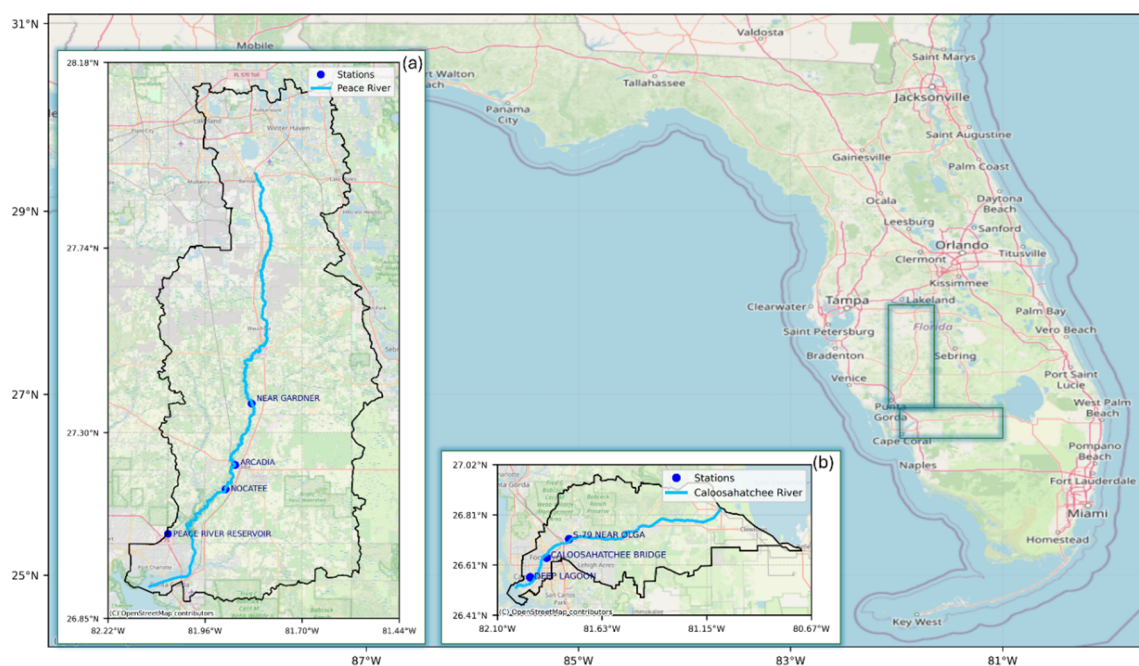


Figure 2. Watershed monitoring networks and hydrological boundaries. Delineation of the (a) Peace River and (b) Caloosahatchee River watersheds using HUC-8 boundaries. Blue markers indicate the USGS and SFWMD gauging stations (e.g., S-79 structure) used to quantify freshwater discharge, total nitrogen (TN), and total phosphorus (TP) loading into the Charlotte Harbor and San Carlos Bay systems.

2.4. Combined Environmental Dataset

All biological and environmental variables were integrated into a single, temporally aligned dataset to serve as the feature matrix for machine learning. *K. brevis* cell counts, salinity, and water temperature were aggregated into weekly intervals using maximum values to preserve physiological and environmental extremes (Brand & Compton, 2007). Physical drivers, including wind vector components and sea surface height (SSH) anomalies, were aggregated into weekly means to represent persistent forcing conditions. Environmental datasets associated with the Peace River basin including river discharge and nutrient concentrations were merged with oceanographic and atmospheric variables on a common weekly time index. Corresponding weekly discharge and nutrient data for the Caloosahatchee River were processed separately and subsequently joined to the Peace River dataset using an outer join on the time variable. This approach preserved the full temporal coverage of both river systems while allowing for asymmetric data availability across watersheds. Following dataset integration, rows containing missing values across predictor variables were removed to ensure a complete feature matrix for model training and evaluation. This listwise deletion was applied after all lagged predictors were constructed to avoid partial feature vectors. The final integrated dataset for machine learning training and validation, providing a multi-decadal record of biological responses and regional drivers, is summarized in Figure 3 and detailed in the supplementary materials (Duus, 2026; A. S. Elshall, 2025).

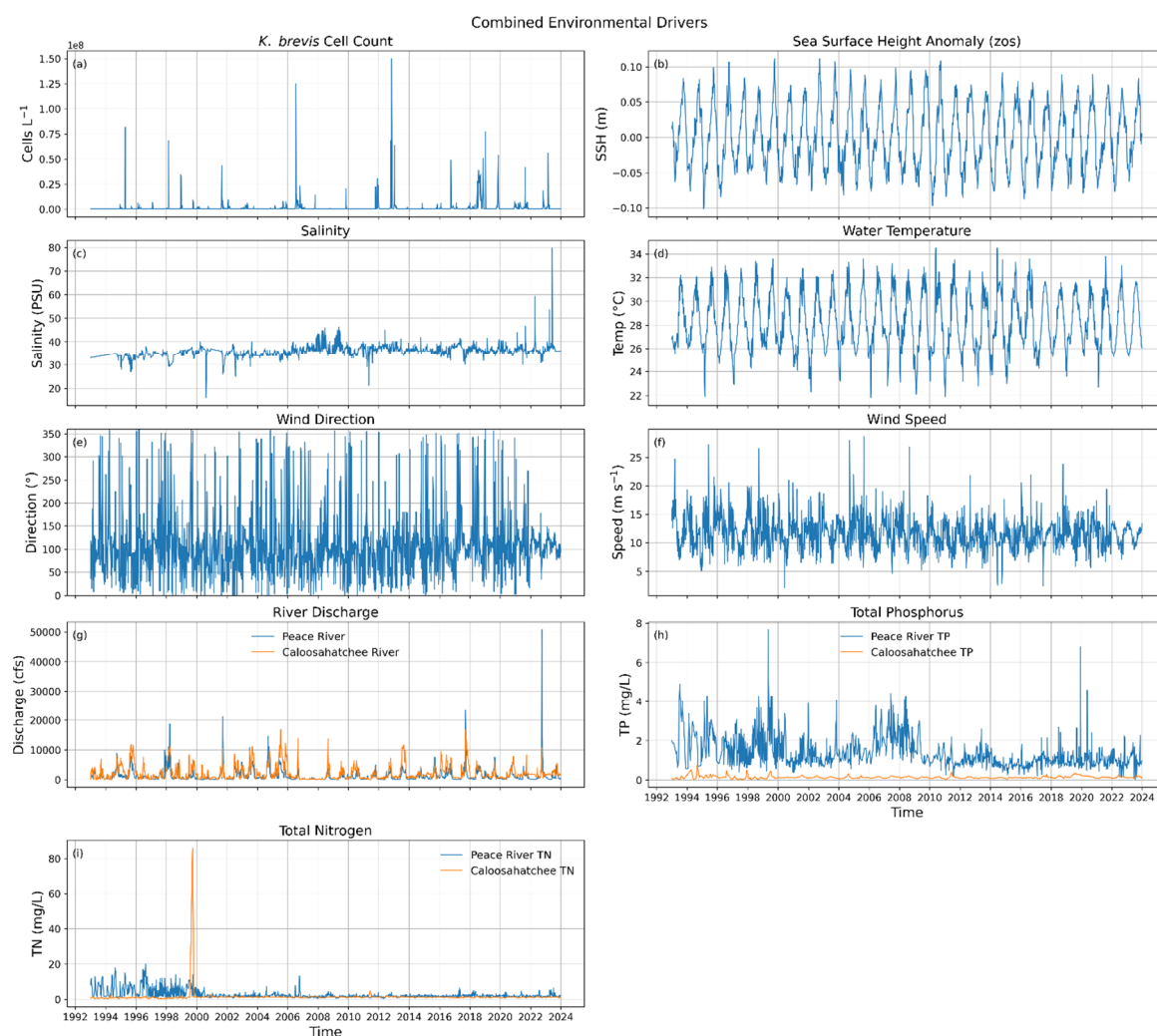


Figure 3. Integrated weekly feature matrix for machine learning model development. Multi-panel time series illustrating the combined environmental dataset spanning 1990–2024. Panels show (a) *K. brevis* cell counts, (b) sea surface height anomaly, (c) salinity, (d) water temperature, (e) wind direction, (f) wind speed, (g) river discharge for the Peace and Caloosahatchee Rivers, (h) total phosphorus, and (i) total nitrogen. Together, these

time-aligned biological, physical, and biogeochemical variables form the integrated predictor set used for Random Forest model training and validation.

2.5. Machine Learning Framework

K. brevis bloom occurrence was treated as a binary classification task. Model predictions target weekly bloom occurrence one week ahead ($t + 1$), using predictor information available at week t , including contemporaneous environmental variables and lagged terms. A threshold of $\geq 100,000$ cells L^{-1} was applied to define the "bloom" class, with all other values designated as "non-bloom" (FWC-FWRI, 2025b). To evaluate the influence of expanded spatial sampling on model performance, the *K. brevis* response record was expanded to include coastal bloom observations from the Tampa Bay region. Tampa Bay observations were merged into the dataset as an additional bloom observation time series; however, model targets were defined using the Southwest Florida *K. brevis* record (kb) and the integrated environmental predictors. This augmentation increased spatial coverage of bloom occurrence and improved temporal continuity of the concentration record, while retaining the same environmental predictor framework. The dataset was split chronologically to preserve temporal structure and prevent information leakage. All observations prior to 1 January 2019 were used for model training, while a contiguous holdout period from 2019 onward was reserved exclusively for testing. Each sample represents a single weekly observation (one row per week) formed by merging regionally compiled bloom observations and environmental predictors on a common weekly time index. No random shuffling was applied. Feature scaling was performed using a RobustScaler fit only on the training data and subsequently applied to the test data.

The primary predictive model was a Random Forest (RF) classifier, selected for its ability to handle non-linear interactions and its robustness against overfitting in imbalanced datasets (Li et al., 2021). To account for biological persistence and the local retention of biomass, lagged predictors (one- and two-week intervals) were constructed for *K. brevis* cell counts and watershed inputs (Yan et al., 2024). To address class imbalance (the prevalence of non-bloom weeks over bloom weeks), class-weighted model fitting was implemented to ensure the minority bloom class contributed proportionally during the training phase (Li et al., 2021). Predictor variables were scaled using a normalization approach based on the median and interquartile range to reduce sensitivity to hydrological and nutrient extremes as detailed in Duus (2026). The supervised classification model was evaluated within a common preprocessing framework. Models were implemented using established open-source machine learning libraries including scikit-learn using default hyperparameter settings (Pedregosa et al., 2011).

2.6. Model Evaluation and Interpretability

Model performance was evaluated on a withheld test subset using metrics optimized for imbalanced environmental data. These included balanced accuracy, which averages sensitivity and specificity, and the F1-score, which provides a harmonic mean of precision and recall (Li et al., 2021; Medina et al., 2024). Precision quantifies the proportion of predicted bloom events that were correctly identified, reflecting the model's ability to minimize false-positive bloom predictions. Recall measures the proportion of observed bloom events that were correctly detected, indicating the model's sensitivity to bloom occurrence. The F1-score therefore provides a balanced measure of bloom prediction performance under class imbalance. The receiver operating characteristic area under the curve (ROC-AUC) evaluates the model's ability to distinguish between bloom and non-bloom conditions across the full range of probability thresholds used to convert continuous model outputs into binary classifications. Balanced accuracy computes the mean of sensitivity (how well the model detects bloom events) and specificity (how well the model correctly identifies non-bloom conditions), thereby accounting for unequal class frequencies and providing an unbiased measure of model performance for both bloom and non-bloom conditions. Model interpretability was assessed using permutation-based feature importance, quantifying the sensitivity of the model's predictive skill to individual environmental drivers implemented using standard routines in scikit-learn

(Pedregosa et al., 2011). Partial dependence plots were utilized to visualize the marginal effects of river discharge and nutrient concentrations on bloom probability, as conditional model responses rather than mechanistic thresholds.

3. Results and Discussion

3.1. Characterization of Regional Environmental Drivers

The integrated multi-decadal record (1990–2024) of environmental predictors and biological responses is illustrated in Figure 3, which highlights the temporal alignment of *K. brevis* cell counts with physical and biogeochemical forcing. The compiled dataset reveals episodic, high-magnitude nutrient loading and river discharge events that are characteristic of the southwest Florida coastal system (Medina et al., 2022). These pulses are particularly evident in the late 1990s and during active hurricane seasons. This reflects the preservation of short-duration loading events that have been shown to exert a disproportionate influence on coastal phytoplankton dynamics (Heil, Dixon, et al., 2014; Glibert et al. 2026). Hydrographic conditions recorded concurrently with biological samples demonstrate clear seasonal variability, as shown in the supplementary material (A. S. Elshall, 2025). Salinity minima typically coincide with seasonal rainfall and high river discharge from the Peace and Caloosahatchee Rivers, which deliver elevated nitrogen and phosphorus loads to the nearshore environment (Brand & Compton, 2007; Medina et al., 2022). Conversely, peak *K. brevis* cell counts often occur during periods of elevated water temperatures in late summer and fall, consistent with the physiological optima for the species (Heil, Dixon, et al., 2014; Stumpf et al., 2022).

Physical forcing on the West Florida Shelf is further characterized by wind and circulation patterns. The wind rose in Figure 4 illustrates the distribution of wind vectors at NOAA buoy 42003, revealing a dominance of northeasterly and southeasterly components that reflect the regional seasonal atmospheric variability on the shelf. Northerly wind components facilitate coastal upwelling and are associated with shoreward transport of blooms initiated offshore (Maze et al., 2015; Weisberg et al., 2014). Furthermore, westerly (onshore) wind components play a vital role in bloom retention by preventing algal biomass from dispersing offshore and instead trapping nutrient-rich water against the coastline (Li et al., 2021; Stumpf et al., 2022). This shoreward retention can maintain elevated nearshore cell concentrations and is associated with increased respiratory irritation risk, as onshore winds facilitate the transport of aerosolized brevetoxins onto inhabited beaches (Stumpf et al., 2022). Conversely, the frequent offshore easterly winds observed in the wind rose often correspond with reduced coastal exposure, transporting surface blooms away from the coast and reducing public health impacts even when high biomass is present in the study area (Stumpf et al., 2022; Weisberg et al., 2014). Large-scale circulation, represented by sea surface height (SSH) anomalies capture the variability of the Loop Current. Positive SSH anomalies reflect the northward penetration of the Loop Current, a physical configuration that has been associated with large-scale red tide manifestation along the southwest Florida coast (Elshall et al., 2022; Maze et al., 2015; Weisberg et al., 2014).

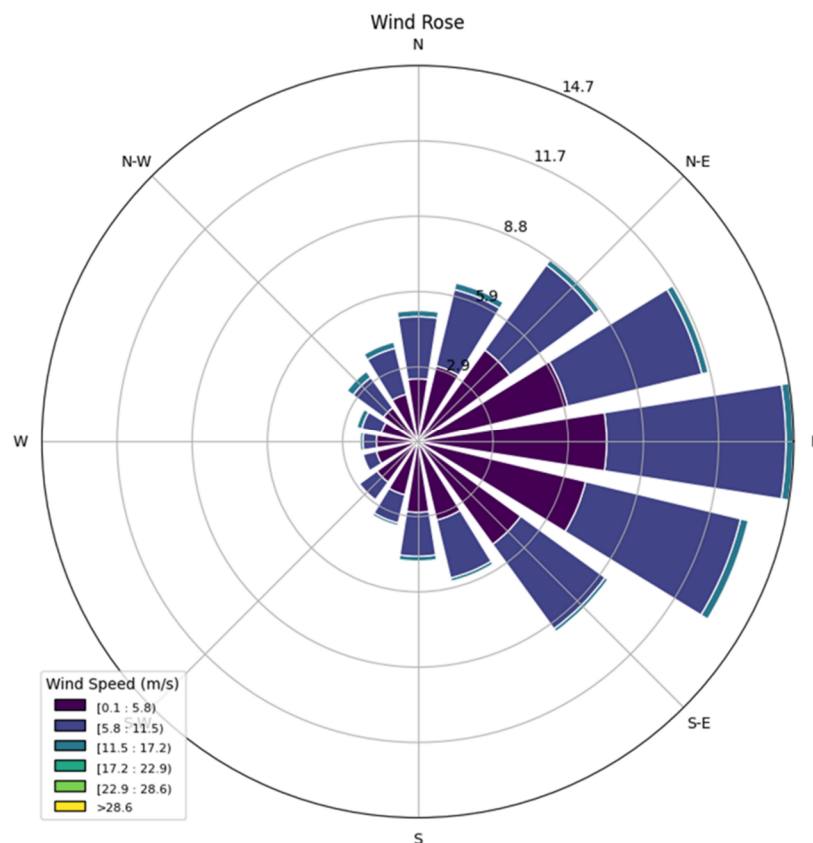


Figure 4. Wind rose showing the frequency distribution of wind speed and direction at NOAA buoy 42003. The wind field is dominated by easterly to southeasterly winds, indicating frequent offshore-directed flow on the West Florida Shelf. Episodic northerly and westerly wind events, although less frequent, are important for regulating upwelling-driven transport and the nearshore retention of algal biomass.

3.2. Model Predictive Performance

The Random Forest classifier demonstrated strong skill in predicting *K. brevis* bloom occurrence, achieving an overall accuracy of 90% and a balanced accuracy of 87.6%. Reliable performance across both bloom and non-bloom classes is a critical requirement for this framework, given the inherent class imbalance found in multi-decadal biological monitoring records (Li et al., 2021; Medina et al., 2024). For the minority bloom class ($\geq 100,000$ cells L^{-1}) the model reached an F1-score of 0.85, supported by high precision (90%) and recall (88%). As illustrated in the Confusion Matrix (Figure 5), the model correctly identified 165 out of 172 true non-bloom weeks and 69 out of 87 true bloom weeks, indicating an ability to distinguish between bloom and non-bloom conditions while minimizing both missed events and false alarms.

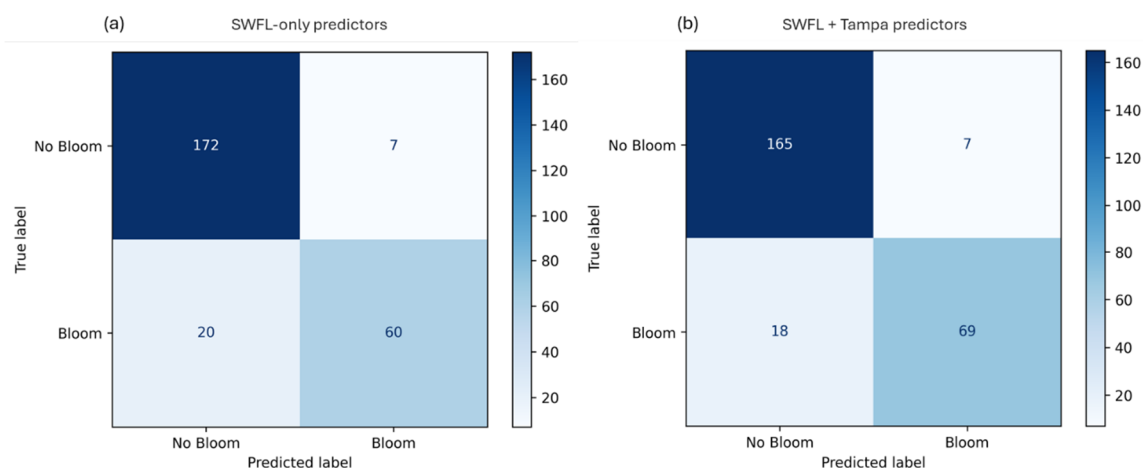


Figure 5. Confusion matrices for Random Forest bloom classification models evaluated on a withheld test set. Panel (a) shows results using the Southwest Florida (*K. brevis*) response record, while panel (b) reflects model evaluation following inclusion of additional Tampa Bay bloom observations in the merged dataset. In both cases, non-bloom conditions are classified with high accuracy and a low false-positive rate (7 events), indicating conservative bloom declarations. When additional bloom observations are available, bloom detection improves, increasing true positive classifications from 60 to 69 and reducing false negatives from 20 to 18. Correspondingly, bloom recall increases from 75% to 79%, and balanced accuracy improves from 85.5% to 87.6%, demonstrating enhanced sensitivity to bloom conditions without increasing false alarms.

Model generalization and the impact of data density were evaluated using the learning curve (Figure 6). Validation accuracy increased steadily with training set size before plateauing near 84%. While the model achieved a training accuracy of 1.00 across all sample sizes, reflecting strong memorization capacity, the persistent generalization gap between training and validation scores suggests a tendency toward overfitting. Exposure to a broader range of bloom and non-bloom conditions through additional observations was associated with improved model sensitivity, though the leveling of validation accuracy near 84% indicates that predictive performance remains constrained by data availability and variability. Historical bloom records are inherently imbalanced, with substantially more non-bloom weeks than bloom weeks. Although class weighting and feature engineering helped address this imbalance, further improvement may be achieved by incorporating additional observations or satellite-derived bloom proxies to increase the density of bloom-related data (Ai et al., 2023; Park et al., 2024).

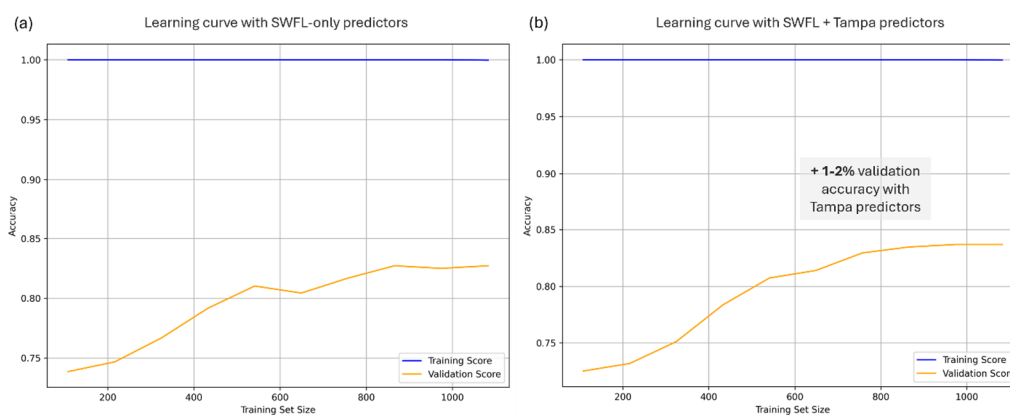


Figure 6. Learning curves for Random Forest bloom classifiers trained with different predictor sets. Learning curves compare models trained using SWFL-only predictors with an expanded dataset including additional Tampa Bay bloom observations. Validation accuracy increases with training set size in both cases and plateaus near ~84%, indicating limits to model generalization imposed by variability in the multi-decadal record. The

expanded dataset provides a modest but consistent improvement in validation accuracy ($\approx 1\text{--}2\%$), indicating enhanced sensitivity without altering the overall performance ceiling.

The stability of the classification threshold was further assessed using the precision–recall curve (Figure 7). The classifier maintained strong precision (~ 0.90) as recall increased toward ~ 0.80 , with a notable performance drop occurring only at the highest recall levels (>0.90). This trade-off is characteristic of imbalanced environmental classification problems, where aggressively attempting to capture all bloom events increases false-positive predictions (Li et al., 2021). Overall, the model's ability to capture the timing and persistence of bloom activity with high consistency demonstrates the effectiveness of the regionally integrated machine learning framework for short-term bloom occurrence prediction.

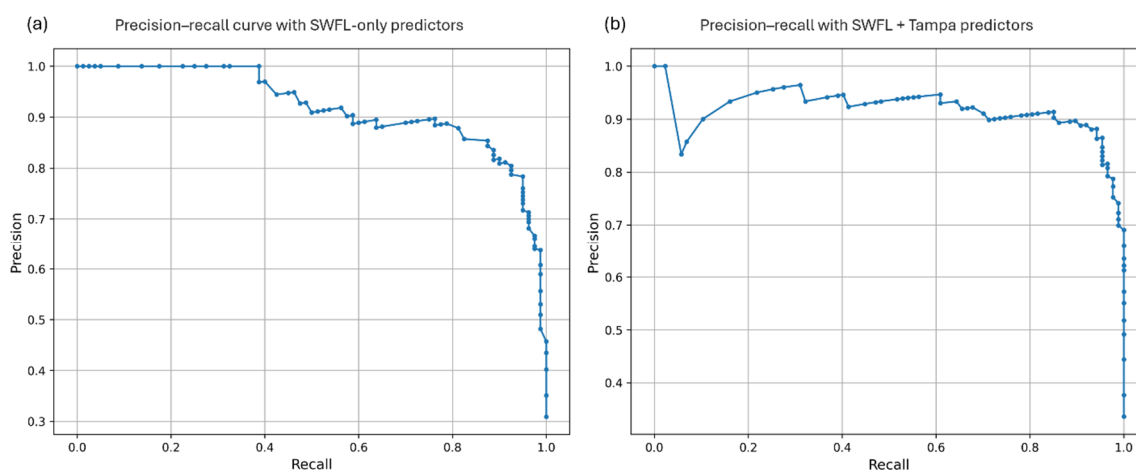


Figure 7. Precision–recall curves for Random Forest bloom classification models using different predictor sets. Curves are shown for models trained with (a) SWFL-only predictors and (b) an expanded dataset including additional Tampa Bay bloom observations. In both cases, high precision is maintained across a broad range of recall values, with a more pronounced trade-off between sensitivity and false-positive rates emerging beyond a recall of approximately 0.80. The expanded dataset sustains high precision over a wider recall range, indicating improved discrimination without a substantial increase in false positives.

A distinguishing aspect of this study is the integration of nutrient data from both the Peace and Caloosahatchee Rivers into a single predictive framework. Previous analyses often focused on a single watershed, whereas this approach captures a broader representation of nutrient dynamics across southwest Florida. By incorporating data from both rivers along with bloom observations from Tampa Bay, the model reflects the combined influence of multiple freshwater sources. This regional scope may have contributed to improved predictive robustness compared to localized models (Ananias et al., 2022; Yan et al., 2024). The findings suggest that predictive performance benefits from a multi-watershed perspective that captures cumulative nutrient forcing. Although this study focuses on southwest Florida, the approach could be adapted to other regions with complex freshwater–coastal interactions, such as estuaries or river plume systems, provided that sufficient historical records are available (Li et al., 2021; Wells et al., 2015).

3.3. Biological Persistence and Influence of Lagged Features

A significant factor in the strong predictive skill of the Random Forest model was the inclusion of lagged biological predictors, specifically *K. brevis* cell concentrations from the preceding one and two weeks. As shown in Figure 8, the model's predicted bloom status closely tracks the timing and duration of actual events, showing strong agreement during ongoing bloom periods. This performance during sustained events reflects the fact that coastal blooms do not dissipate instantaneously. Instead, they persist through a combination of biological continuity, local retention

of biomass, and shelf-scale advection. In this context, lagged predictors function as indicators of short-term persistence rather than independent environmental forcing. However, the model's reliance on these autoregressive features can introduce a potential weakness in detecting abrupt bloom emergence, defined as transitions from a non-bloom state to a bloom state following an extended period of absence. As illustrated by the "Actual Versus Predicted" comparison in Figure 8, while the model identifies bloom start dates with minimal lag, its sensitivity may be reduced in scenarios where recent bloom history is absent. This reflects a common challenge in machine learning for HABs, where models may learn to repeat the most recent observation as a shortcut to high accuracy, essentially "catching up" to reality rather than providing a leading indicator of initiation. Thus, while the model excels at short-term operational forecasting, its skill would likely decrease for long-lead initiation forecasts if biological lags were excluded. Accordingly, the current model-structure prioritizes predictive reliability for ongoing public health threats over mechanistic initiation modeling. Future sensitivity experiments that isolate environmental drivers from biological lags are warranted to quantify the relative contribution of external forcing such as nutrient loading and wind-driven transport during the critical onset phase. Despite this limitation, the Random Forest architecture utilized here remains more robust for this data structure than simple linear models, which fail to capture the nonlinear threshold effects inherent in biological persistence.

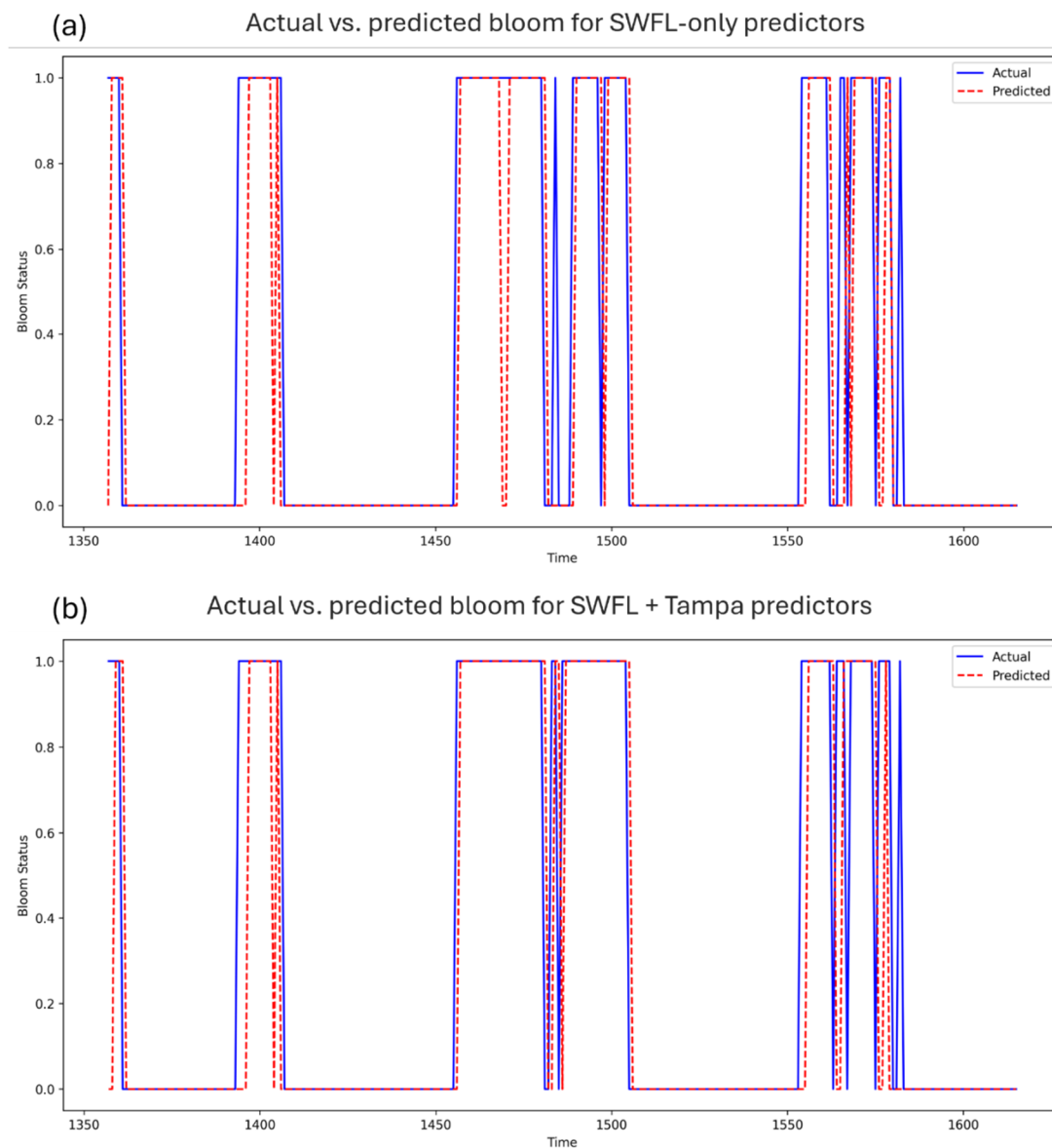


Figure 8. Comparative time series of actual versus predicted bloom status for models using different predictor sets. Observed *K. brevis* bloom occurrence (solid blue) is compared with Random Forest predictions (dashed red) for the test period using (a) SWFL-only predictors and (b) an expanded predictor set including Tampa-derived variables. In both cases, the models capture the persistence and termination of bloom events with high fidelity. The inclusion of Tampa predictors improves temporal alignment during complex bloom periods and reduces intermittent misclassification, while reliance on short-term biological lags continues to limit sensitivity during the initial transition from non-bloom to bloom conditions.

3.4. Synergistic Effects of Nutrient Loading and Discharge

While physical forcing governs the transport and retention of biomass, the intensity and duration of coastal blooms are associated with the availability of nutrients (Medina et al., 2022; Yan et al., 2024). The Random Forest model identified nonlinear relationships between watershed inputs and bloom probability, as illustrated by the partial dependence plots (Figure 9). Contrary to a simple linear dose-response relationship, the model indicates that bloom probability rises sharply with low-to-moderate increases in river discharge and total nitrogen (TN) concentrations before plateauing or slightly declining at the highest values. This plateau suggests a model-inferred saturation range, which is approximately 2 - 3 mg L⁻¹ for TN and 1.5 mg L⁻¹ for Total Phosphorus (TP), beyond which additional loading does not proportionally increase bloom risk. Hydrological conditions, particularly discharge from the Peace and Caloosahatchee Rivers, were also key drivers of bloom probability. The model indicates that moderate discharge events, often following seasonal rainfall or runoff pulses, are associated with elevated bloom probability.

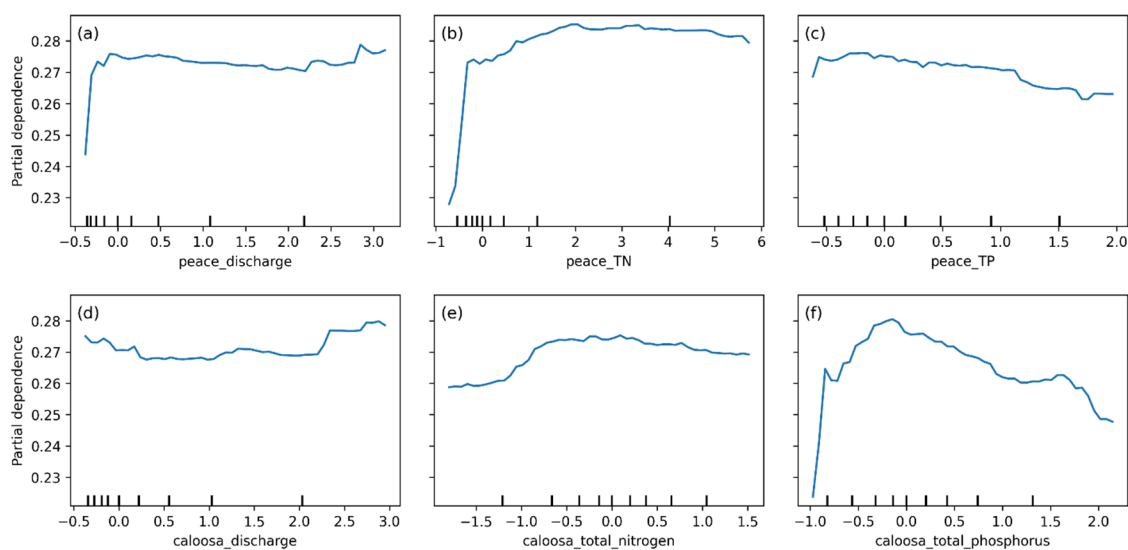


Figure 9. Partial dependence plots for hydrological and nutrient predictors from the final Random Forest model. The marginal effects of (a) Peace River discharge, (b) Peace River total nitrogen (TN), and (c) Peace River total phosphorus (TP), along with (d) Caloosahatchee River discharge, (e) Caloosahatchee River total nitrogen, and (f) Caloosahatchee River total phosphorus, on the predicted probability of *K. brevis* bloom occurrence are shown. The curves illustrate a sharp increase in modeled risk at low-to-moderate levels, followed by a plateau, indicating diminishing marginal sensitivity at extreme values that may reflect dilution or reduced residence time rather than increased bloom support.

Consequently, *K. brevis* blooms were most strongly associated with moderate, rather than extreme, nutrient inputs and river discharge. Pairwise relationships among environmental variables (Figure 10) indicate that blooms tend to occur when TN and TP are elevated but not extreme. This pattern is consistent with nutrient threshold behavior or co-limitation effects (Wang et al., 2016; Yan et al., 2024), where blooms are supported within a narrow range of nutrient availability but are suppressed by very low or excessively high concentrations (Lenes & Heil, 2010; Wells et al., 2015).

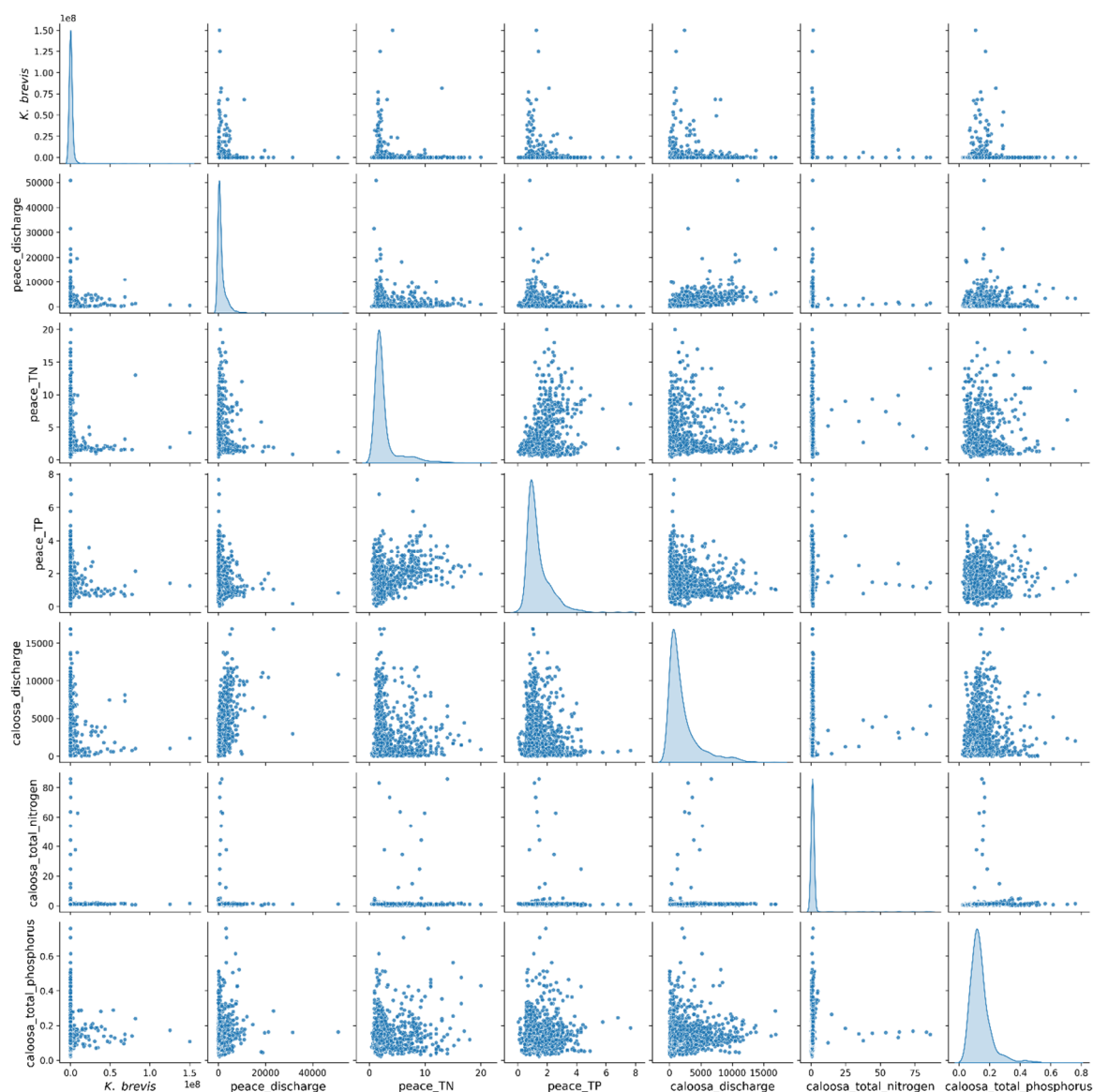


Figure 10. Pairwise relationships among environmental variables. Scatter matrix displaying the joint distributions of *K. brevis* cell counts, river discharge, and nutrient concentrations. The clustering of high cell counts against moderate discharge and nutrient values reinforces the nonlinear nature of bloom dynamics and the rarity of extreme nutrient events in the long-term record.

The interactions between specific nutrients were further examined using probability contour plots (Figure 11), which indicate a synergistic association between nitrogen and phosphorus availability. Bloom probability remains relatively low when either TN or TP concentrations are minimal but increases substantially when both are elevated simultaneously. Specifically, the highest modeled probabilities (>0.6) occur when TN exceeds $\sim 8\text{--}10\text{ mg L}^{-1}$ and TP exceeds $\sim 1.5\text{ mg L}^{-1}$. These results are consistent with prior evidence that *K. brevis* blooms on the West Florida Shelf may reflect balanced nutrient stoichiometry and co-limitation rather than control by a single limiting nutrient (Heil, Dixon, et al., 2014; Lenés & Heil, 2010; Medina et al., 2022; Vargo et al., 2008, 2008; Wang et al., 2016; Wells et al., 2015).

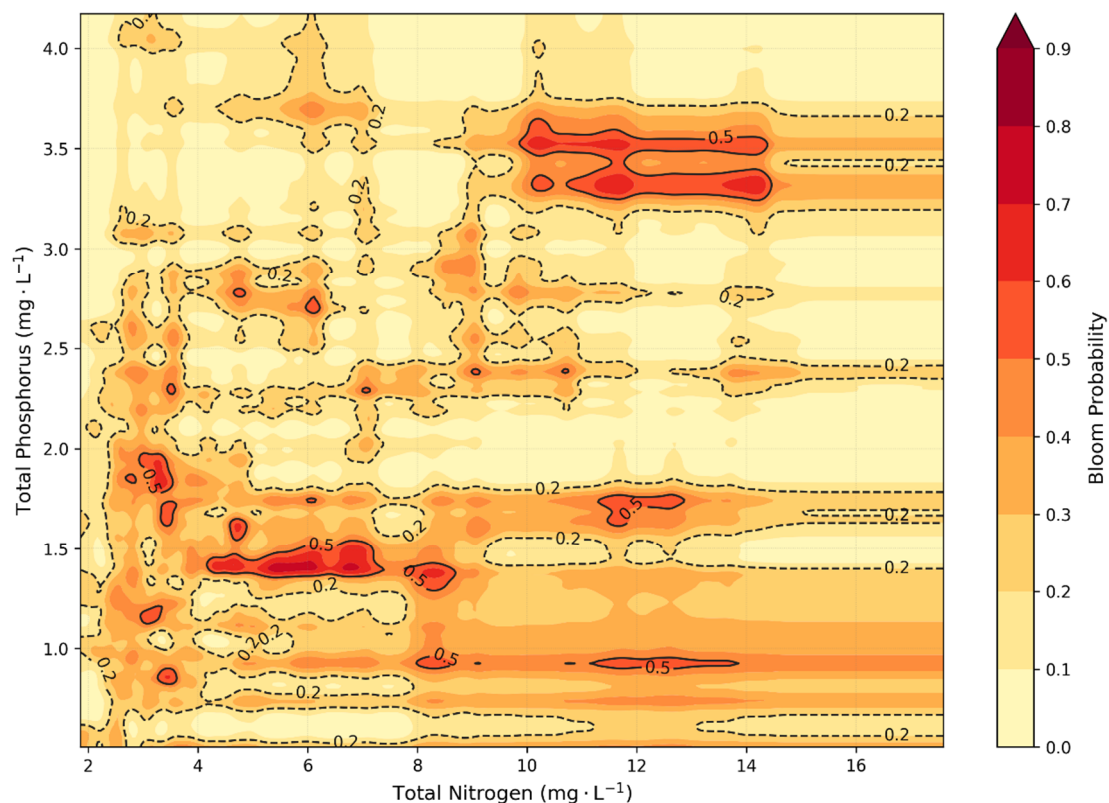


Figure 11. Two-dimensional probability contours of nutrient interactions. Bloom probability mapped as a function of concurrent total nitrogen (TN) and total phosphorus (TP) concentrations using partial dependence from the Random Forest model. The color gradient (light to dark red) indicates increasing modeled likelihood of bloom occurrence. The color gradient (light yellow to deep red) indicates the modeled likelihood of a *K. brevis* bloom ($\geq 100,000$). Isopleths denote specific probability thresholds, including low-risk (dashed line, $P = 0.2$) and moderate-to-high risk (solid line, $P \geq 0.5$). The smooth topography results from the application of a Gaussian filter ($\sigma = 2.0$), providing a continuous representation of the probability density while removing model-induced aliasing artifacts. Regions of elevated probability emerge when both nutrients are concurrently elevated, consistent with a synergistic association, while irregular contour structure reflects data sparsity and nonlinear model responses rather than sharp ecological thresholds.

The lack of a consistent correspondence between extreme discharge events and bloom occurrence suggests that freshwater inflow may sometimes dilute or disperse blooms rather than intensify them (Weisberg et al., 2016). Large discharge events can also lower nearshore salinity below levels favorable for *K. brevis* growth (~ 24 ppt), further limiting bloom development (Steidinger et al., 1998). This nonlinearity reflects the complex interaction between hydrology, salinity, and nutrient delivery, and indicates that river inputs play a dual role in bloom dynamics depending on timing and magnitude. For example, while riverine discharge delivers essential nutrients, extreme flow events can reduce the local residence time necessary for algal accumulation, effectively flushing biomass from the estuary (Phlips et al., 2023; Weisberg et al., 2016). Potential alternate explanations for the observed relationships can be considered. For example, correlations between river discharge and bloom events may reflect broader seasonal patterns rather than direct nutrient effects (Roelke & Pierce, 2011). Offshore nutrient sources, such as nitrogen fixation or upwelling, were not accounted for, and these processes could play a role in sustaining blooms during certain seasons (Lenes & Heil, 2010; Weisberg et al., 2016). Future models that integrate offshore nutrient data may provide a more complete picture of bloom dynamics.

However, caution is required when interpreting these nutrient-bloom associations. As shown in the pairwise relationship plot (Figure 11), the distributions of nutrient data are strongly right-skewed, meaning that extreme high-nutrient events are rare in the training record. Recent analyses have

reported positive relationships between Caloosahatchee River TN loading and *K. brevis* bloom severity or duration across both long-term records and recent bloom periods (Glibert et al., 2025; Tomasko et al., 2024) suggesting that increased watershed nitrogen inputs can enhance bloom persistence once populations are established nearshore. In contrast, the present machine-learning analysis indicates a nonlinear response, with peak bloom probabilities occurring at moderate nutrient levels, while predictive uncertainty increases under extreme discharge conditions that are sparsely represented in the dataset, resulting in wider confidence intervals at these upper ranges. Furthermore, while the model successfully uses nearshore nutrient concentrations as predictors, this does not confirm a strictly causal "bottom-up" driving mechanism. Elevated nutrient concentrations within a bloom may be partly a result of the bloom itself, generated through biological regeneration, zooplankton excretion, and the decay of fish kills, rather than solely representing external watershed loading (Heil, Dixon, et al., 2014; Killberg-Thoreson et al., 2014; Walsh et al., 2009). Therefore, while the machine learning framework confirms that watershed discharges are strong predictors of bloom maintenance and intensification in the nearshore environment (Medina et al., 2022), they should not be interpreted as primary drivers of offshore bloom initiation (K. A. Steidinger, 2009; Weisberg et al., 2019).

3.5. Role of Physical Forcing

While biological persistence and nutrient availability emerged as the dominant predictors in the Random Forest model, physical oceanographic variables—specifically wind forcing and sea surface height (SSH) anomalies—retained moderate importance, acting as essential modulators of bloom transport and maintenance. As indicated in the feature importance ranking (Figure 12), wind speed and direction were secondary to riverine inputs but consistently contributed to model accuracy. This ranking aligns with the prevailing bloom hypothesis, where physical forcing is not the primary source of biomass generation but is the critical mechanism for delivering offshore-initiated populations to the nearshore environment (K. A. Steidinger, 2009; Walsh et al., 2006). The model's sensitivity to wind direction reflects the physical necessity of onshore and downwelling-favorable winds for concentrating *K. brevis* cells against the coast. Periods of onshore winds can retain nutrient-rich water nearshore, creating conditions favorable for bloom maintenance (Basterretxea et al., 2024; Pitcher et al., 2010). Analysis of the wind rose (Figure 4) confirms that easterly winds are the prevailing wind direction on the West Florida Shelf. However, model sensitivity analysis indicates that less frequent westerly and southwesterly wind events are associated with accumulation of surface populations in the surf zone, a process documented to intensify respiratory irritation events (Stumpf et al., 2022). Conversely, persistent easterly (offshore-directed) winds tend to disperse surface blooms seaward, reducing nearshore severity even if the bloom persists offshore (Maze et al., 2015; Weisberg et al., 2016). Future work could examine these atmospheric effects more directly, particularly their interaction with coastal circulation and transport.

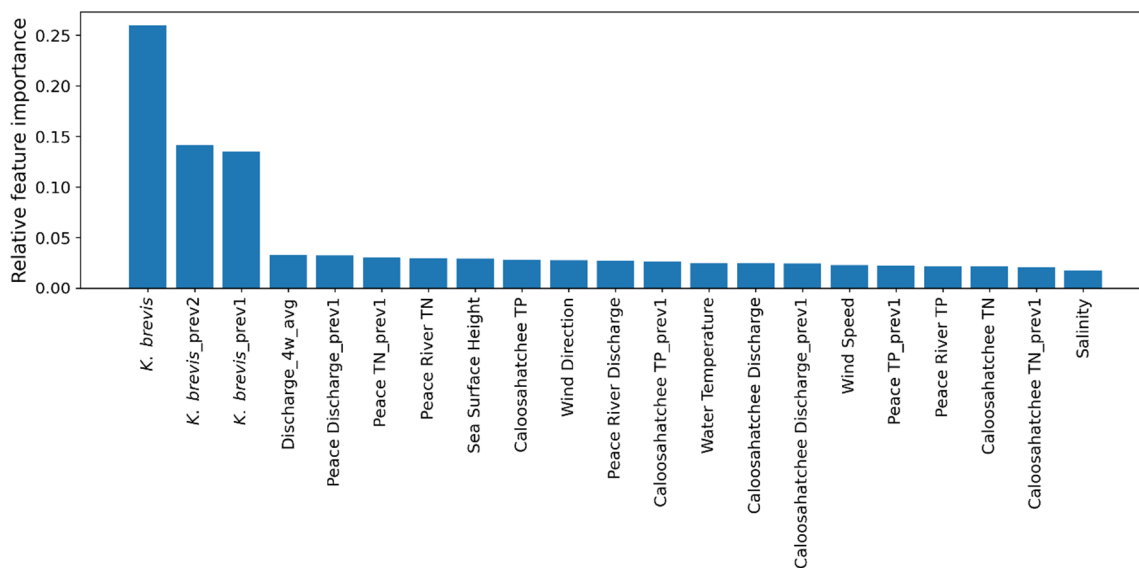


Figure 12. Feature importance ranking for the Random Forest model. The relative contribution of each predictor variable to model accuracy. *K. brevis* autoregressive terms (current and lagged cell counts) and Peace River discharge dominate the ranking, reflecting the importance of biological persistence and watershed nutrient pulses. Physical variables (SSH, wind speed, temperature) play a secondary but non-negligible role in modulating bloom probability.

The influence of the Loop Current, represented by SSH anomalies, was captured by the model but ranked lower than watershed variables (Figure 12). While positive SSH anomalies (indicative of a northward Loop Current extension) have been statistically associated with increased retention of waters on the West Florida Shelf, a condition favorable for bloom maintenance (Maze et al., 2015), the modest influence of SSH within the model indicates that complex shelf interactions controlling bloom persistence may not be fully resolved using surface elevation alone. The machine learning model may struggle to resolve the specific "pressure point" interactions at the shelf slope (Weisberg et al., 2014, 2016). These interactions drive deep-layer upwelling that can either fuel blooms or, in cases of extreme duration (e.g., 2010), suppress them by flushing the shelf with excess inorganic nutrients that favor faster-growing diatoms over *K. brevis* (Weisberg et al., 2014, 2016).

The model exhibited weaker sensitivity to SST than expected, given that temperature can affect *K. brevis* growth rates (Wells et al., 2015). One explanation is that SST shows relatively minor seasonal fluctuations in the Gulf compared to nutrient and discharge dynamics, making it a less prominent predictor on a weekly scale. Alternatively, temperature effects may be indirectly captured through correlated variables such as seasonal rainfall or runoff (Pitcher et al., 2010), or may act primarily as a background environmental condition that modulates bloom potential rather than serving as a direct short-term predictor (Glibert et al., 2025). Similarly, salinity, while important in bloom ecology, did not dominate the predictive signal, likely due to its strong correlation with discharge events already accounted for in the dataset (Weisberg et al., 2016).

3.6. Model Limitations

A notable limitation is that the model uses surface nutrient concentrations as proxies for bloom potential, while subsurface processes such as upwelling and benthic nutrient fluxes were not represented (Pitcher et al., 2010; Weisberg et al., 2016). Biological factors, such as grazing pressure and competition with other phytoplankton, were also excluded. These processes can influence bloom development and persistence, particularly when environmental conditions are borderline favorable (Griffith & Gobler, 2020; Wells et al., 2015). Adding biological or optical indicators such as chlorophyll-a concentrations or particulate matter data could improve predictive performance in

future iterations (Ai et al., 2023; Park et al., 2024; Song, 2025). Furthermore, the use of weekly aggregated wind data may smooth out rapid, synoptic-scale wind events (e.g., frontal passages) that drive immediate biological responses. While the model effectively forecasts the maintenance of existing nearshore blooms based on nutrient and autoregressive signals, its ability to predict the initial onshore arrival of a bloom remains constrained by the lack of subsurface data. Future iterations incorporating outputs from hydrodynamic models (e.g., WFCOM) or subsurface glider data as features could significantly bridge this gap between surface proxies and 3D physical reality (Vargo et al., 2008; Weisberg et al., 2016).

The Random Forest approach provided high accuracy and interpretability but does not explicitly model temporal dependencies between consecutive weeks. Instead, temporal structure was represented through the inclusion of lagged predictors, particularly recent *K. brevis* concentrations, which capture short-term biological persistence and local retention or redistribution of bloom biomass in coastal waters. Similar persistence-based effects have been noted in previous HAB modeling studies, where recent bloom history strongly influences near-term bloom probability (Ai et al., 2023; Yan et al., 2024). While this strategy improves predictive skill during ongoing bloom conditions, it may reduce sensitivity to true bloom emergence events, defined as abrupt transitions from extended non-bloom periods to bloom conditions. In such onset scenarios, where bloom-history predictors are weak or absent, model predictions rely more heavily on external environmental forcing, including nutrient loading, river discharge, wind-driven transport, and ocean circulation variability. This limitation is consistent with prior findings that tree-based and autoregressive ML models tend to favor persistence over initiation dynamics and may struggle to capture abrupt bloom onset without explicit temporal sequence modeling (Huang et al., 2025; Yan et al., 2024). Time-series-oriented approaches, such as recurrent neural networks or hybrid frameworks that explicitly encode temporal evolution, have therefore been proposed as complementary tools for improving bloom forecasting and the detection of emerging bloom conditions (Ai et al., 2023; Lin et al., 2023). Future work should explicitly evaluate model performance with and without bloom-history predictors to quantify the relative contributions of persistence versus external environmental forcing, particularly for forecasting bloom emergence following extended non-bloom periods. Finally, while the model performed well on historical data, applying it operationally would require continuous updates to account for evolving climate and land-use patterns (Ralston & Moore, 2020; Wells et al., 2015). Changes in rainfall intensity, hurricane frequency, or sea-level rise could alter the historical relationships on which the model is based (Griffith & Gobler, 2020). Periodic recalibration using recent data and scenario-based simulations could help ensure the model remains robust under changing environmental conditions.

4. Conclusions

This study demonstrates that *K. brevis* bloom forecasting on the West Florida Shelf is enhanced by integrating multi-watershed nutrient and hydrological data into a machine learning framework. The Random Forest classifier achieved robust predictive performance, determining bloom occurrence with a balanced accuracy of 88.7% and an F1-score of 0.86. The model displayed high precision of 0.90, indicating a low rate of false alarms, which is a critical requirement for operational public health advisories and coastal resource management (Stumpf et al., 2022). Together, these results demonstrate that data-driven approaches can reliably discriminate bloom and non-bloom conditions despite strong class imbalance and environmental variability.

The application of interpretable machine learning techniques, specifically partial dependence plots and feature importance rankings, revealed nonlinear relationships between environmental drivers and bloom dynamics. Contrary to simple linear dose-response assumptions, bloom probability increased sharply with low-to-moderate increases in river discharge and nutrient loading before reaching a saturation plateau. This pattern suggests that while riverine inputs are essential for bloom maintenance, extreme discharge events may produce contrasting outcomes, either enhancing bloom persistence through increased nutrient loading (Glibert et al., 2025; Tomasko et al., 2024) or

yield diminishing returns or potentially dispersive effects due to reduced residence times (Phlips et al., 2023; Weisberg et al., 2016). Furthermore, the analysis identified synergistic interactions between total nitrogen (TN) and total phosphorus (TP), supporting the hypothesis that coastal blooms are co-limited and that dual-nutrient management strategies are necessary to effectively mitigate bloom intensity (Heil, Dixon, et al., 2014; Medina et al., 2022). While biological persistence, represented by lagged *K. brevis* cell counts, emerged as the dominant predictor for short-term forecasting, physical oceanographic factors provided essential modulation of bloom risk. Sea surface height anomalies and wind forcing were identified as key secondary drivers, consistent with the physical requirements for shoreward transport and shelf retention of biomass initiated offshore (Maze et al., 2015; Weisberg et al., 2019). These results reinforce the coupled biological–physical nature of bloom dynamics on the West Florida Shelf. However, the model’s reliance on surface-based proxies constitutes a limitation, as it likely underrepresents subsurface initiation processes driven by bottom-layer Ekman transport (Weisberg et al., 2016).

This framework provides a practical and interpretable tool for assessing the cumulative influence of the Peace and Caloosahatchee River watersheds on coastal water quality and bloom persistence. To advance from short-term operational forecasting toward improved prediction of bloom emergence, future work must address the generalization gap observed between training and validation performance. This may be achieved through the integration of subsurface hydrodynamic data and the exploration of deep learning architectures. Despite these limitations, the current model offers actionable insights by defining specific hydrological and nutrient conditions associated with elevated red tide severity under evolving climatic and watershed forcing.

Funding statement: This work is funded by U.S. National Science Foundation (NSF) Award Numbers 2536218 and 2536219.

Data Availability Statement: Environmental driver dataset and codes supporting are available at Elshall (2025) and can be directly access from this Jupyter Book (<https://aselshall.github.io/redtides/intro.html>); machine learning codes are available a Duus (2026) and can be directly accessed from this Jupyter Book (<https://mkduus.github.io/red-tide-book/>).

Acknowledgments: The authors thank Carter Baker for his contributions during the early phases of this research, particularly for his work on discharge data processing and initial coding support related to river flow integration.

AI assistance statement: Coding for data analysis and plotting were performed using Python with assistance from GPT-4o. GPT-4o and Gemini 2.5 Pro were used to provided review comments, and to improve text clarity, succinctness, logical flow, and overall polish. AI contributions were verified for accuracy and relevance.

Conflicts of Interest: None.

References

1. Ai, H., Zhang, K., Sun, J., & Zhang, H. (2023). Short-term Lake Erie algal bloom prediction by classification and regression models. *WATER RESEARCH*, 232, 119710. <https://doi.org/10.1016/j.watres.2023.119710>
2. Ananias, P. H. M., Negri, R. G., Dias, M. A., Silva, E. A., & Casaca, W. (2022). A Fully Unsupervised Machine Learning Framework for Algal Bloom Forecasting in Inland Waters Using MODIS Time Series and Climatic Products. *Remote Sensing*, 14(17), Article 17. <https://doi.org/10.3390/rs14174283>
3. Basterretxea, G., Font-Muñoz, J. S., Kane, M., Regaudie-de-Gioux, A., Satta, C. T., & Tuval, I. (2024). Pulsed wind-driven control of phytoplankton biomass at a groundwater-enriched nearshore environment. *Science of The Total Environment*, 955, 177123. <https://doi.org/10.1016/j.scitotenv.2024.177123>
4. Brand, L. E., Campbell, L., & Bresnan, E. (2012). *Karenia*: The biology and ecology of a toxic genus. *Harmful Algae*, 14, 156–178. <https://doi.org/10.1016/j.hal.2011.10.020>
5. Brand, L. E., & Compton, A. (2007). Long-term increase in *Karenia brevis* abundance along the Southwest Florida Coast. *Harmful Algae*, 6(2), 232–252. <https://doi.org/10.1016/j.hal.2006.08.005>

6. Chen, S., & Hu, C. (2017). Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements. *Remote Sensing of Environment*, 201, 115–132. <https://doi.org/10.1016/j.rse.2017.09.004>
7. Dréville, M., Régnier, C., Lellouche, J.-M., Garric, G., & Bricaud, C. (2018). *QUALITY INFORMATION DOCUMENT For Global Ocean Reanalysis Products GLOBAL-REANALYSIS-PHY-001-030*. 48.
8. Duus, M. (2026). *mkduus/red-tide-book: Initial Release for Journal Submission* [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.18165564>
9. Elshall, A. S. (2025). *Machine learning framework for red tide bloom severity classification in Charlotte Harbor, West Florida Shelf*. <https://aselshall.github.io/redtides>
10. Elshall, A. S., Arik, A. D., El-Kadi, A. I., Pierce, S., Ye, M., Burnett, K. M., Wada, C. A., Bremer, L. L., & Chun, G. (2020). Groundwater sustainability: A review of the interactions between science and policy. *Environmental Research Letters*, 15(9), 093004. <https://doi.org/10.1088/1748-9326/ab8e8c>
11. Elshall, A., Ye, M., Kranz, S. A., Harrington, J., Yang, X., Wan, Y., & Maltrud, M. (2022). Earth system models for regional environmental management of red tide: Prospects and limitations of current generation models and next generation development. *Environmental Earth Sciences*, 81(9), 256. <https://doi.org/10.1007/s12665-022-10343-7>
12. Elshall, A., Ye, M., Kranz, S., Harrington, J., Yang, X., Wan, Y., & Maltrud, M. (2021). *Machine learning for red tide prediction in the Gulf of Mexico along the West Florida Shelf*. <https://www.authorea.com/doi/full/10.1002/essoar.10509597?commit=12ed6e15d90fe8993d8ce2cab016c9a737a96629>
13. Fernandez, E., & Lellouche, J. M. (2018). *PRODUCT USER MANUAL For the Global Ocean Physical Reanalysis product GLOBAL_REANALYSIS_PHY_001_030*. 15.
14. Florida Department of Environmental Protection. (2025). *STORET Stations*. <https://geodata.dep.state.fl.us/datasets/storet-stations/about>
15. FWC-FWRI. (2025a). *HAB Monitoring Database*. Florida Fish And Wildlife Conservation Commission. <https://myfwc.com/research/redtide/monitoring/database/>
16. FWC-FWRI. (2025b). *Red Tide Current Status*. Florida Fish And Wildlife Conservation Commission. <https://myfwc.com/research/redtide/statewide/>
17. Glibert, P. M., Heil, C. A., & Li, M. (2025). More sustained, more severe blooms and shifting monthly patterns of the toxigenic dinoflagellate *Karenia brevis* on the West Florida Shelf. *Harmful Algae*, 150, 102967. <https://doi.org/10.1016/j.hal.2025.102967>
18. Griffith, A. W., & Gobler, C. J. (2020). Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems. *Harmful Algae, Climate Change and Harmful Algal Blooms*, 91, 101590. <https://doi.org/10.1016/j.hal.2019.03.008>
19. Heil, C. A., Bronk, D. A., Dixon, L. K., Hitchcock, G. L., Kirkpatrick, G. J., Mulholland, M. R., O'Neil, J. M., Walsh, J. J., Weisberg, R., & Garrett, M. (2014). The Gulf of Mexico ECOHAB: *Karenia* Program 2006–2012. *Harmful Algae, Nutrient Dynamics of Karenia Brevis Red Tide Blooms in the Eastern Gulf of Mexico*, 38, 3–7. <https://doi.org/10.1016/j.hal.2014.07.015>
20. Heil, C. A., Dixon, L. K., Hall, E., Garrett, M., Lenes, J. M., O'Neil, J. M., Walsh, B. M., Bronk, D. A., Killberg-Thoreson, L., Hitchcock, G. L., Meyer, K. A., Mulholland, M. R., Procise, L., Kirkpatrick, G. J., Walsh, J. J., & Weisberg, R. W. (2014). Blooms of *Karenia brevis* (Davis) G. Hansen & Ø. Moestrup on the West Florida Shelf: Nutrient sources and potential management strategies based on a multi-year regional study. *Harmful Algae, Nutrient Dynamics of Karenia Brevis Red Tide Blooms in the Eastern Gulf of Mexico*, 38, 127–140. <https://doi.org/10.1016/j.hal.2014.07.016>
21. Huang, G., Bao, M., Zhang, Z., Gu, D., Liang, L., & Tao, B. (2025). Interpretable Machine Learning-Based Spring Algal Bloom Forecast Model for the Coastal Waters of Zhejiang. *Journal of Ocean University of China*, 24(1), 1–12. <https://doi.org/10.1007/s11802-025-5833-z>
22. Killberg-Thoreson, L., Sipler, R. E., Heil, C. A., Garrett, M. J., Roberts, Q. N., & Bronk, D. A. (2014). Nutrients released from decaying fish support microbial growth in the eastern Gulf of Mexico. *Harmful Algae, Nutrient Dynamics of Karenia Brevis Red Tide Blooms in the Eastern Gulf of Mexico*, 38, 40–49. <https://doi.org/10.1016/j.hal.2014.04.006>

23. Lenes, J. M., & Heil, C. A. (2010). A historical analysis of the potential nutrient supply from the N₂ fixing marine cyanobacterium *Trichodesmium* spp. To *Karenia brevis* blooms in the eastern Gulf of Mexico. *Journal of Plankton Research*, 32(10), 1421–1431. <https://doi.org/10.1093/plankt/fbq061>
24. Li, M. F., Glibert, P. M., & Lyubchich, V. (2021). Machine Learning Classification Algorithms for Predicting *Karenia brevis* Blooms on the West Florida Shelf. *Journal of Marine Science and Engineering*, 9(9), Article 9. <https://doi.org/10.3390/jmse9090999>
25. Lin, S., Pierson, D. C., & Mesman, J. P. (2023). Prediction of algal blooms via data-driven machine learning models: An evaluation using data from a well-monitored mesotrophic lake. *Geoscientific Model Development*, 16(1), 35–46. <https://doi.org/10.5194/gmd-16-35-2023>
26. Marcillo-Yopez, E., Grogan, K. A., Court, C. D., Savchenko, O. M., & Koeneke, R. (2025). Environmental risks and the profitability of Florida's hard clam aquaculture industry. *Aquaculture Economics & Management*, 29(4), 705–740. <https://doi.org/10.1080/13657305.2025.2487523>
27. Maze, G., Olascoaga, M. J., & Brand, L. (2015). Historical analysis of environmental conditions during Florida Red Tide. *Harmful Algae*, 50, 1–7. <https://doi.org/10.1016/j.hal.2015.10.003>
28. Medina, M., Julian, P., Chin, N., & Davis, S. E. (2024). An early-warning forecast model for red tide (*Karenia brevis*) blooms on the southwest coast of Florida. *Harmful Algae*, 139, 102729. <https://doi.org/10.1016/j.hal.2024.102729>
29. Medina, M., Kaplan, D., Milbrandt, E. C., Tomasko, D., Huffaker, R., & Angelini, C. (2022). Nitrogen-enriched discharges from a highly managed watershed intensify red tide (*Karenia brevis*) blooms in southwest Florida. *Science of The Total Environment*, 827, 154149. <https://doi.org/10.1016/j.scitotenv.2022.154149>
30. NDBC. (2025). *NDBC Station Page*. https://www.ndbc.noaa.gov/station_page.php?station=42003
31. 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>
32. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12(null), 2825–2830.
33. Philips, E. J., Badylak, S., Mathews, A. L., Milbrandt, E. C., Montefiore, L. R., Morrison, E. S., Nelson, N., & Stelling, B. (2023). Algal blooms in a river-dominated estuary and nearshore region of Florida, USA: The influence of regulated discharges from water control structures on hydrologic and nutrient conditions. *Hydrobiologia*, 850(20), 4385–4411. <https://doi.org/10.1007/s10750-022-05135-w>
34. Pitcher, G. C., Figueiras, F. G., Hickey, B. M., & Moita, M. T. (2010). The physical oceanography of upwelling systems and the development of harmful algal blooms. *Progress in Oceanography*, 85(1–2), 5–32. <https://doi.org/10.1016/j.pocean.2010.02.002>
35. Ralston, D. K., & Moore, S. K. (2020). Modeling harmful algal blooms in a changing climate. *Harmful Algae, Climate Change and Harmful Algal Blooms*, 91, 101729. <https://doi.org/10.1016/j.hal.2019.101729>
36. Roelke, D. L., & Pierce, R. H. (2011). Effects of inflow on harmful algal blooms: Some considerations. *Journal of Plankton Research*, 33(2), 205–209. <https://doi.org/10.1093/plankt/fbq143>
37. Song, Y. (2025). Forecasting short-term chlorophyll a concentration in Lake Erie using the machine learning XGBoost algorithm. *Environmental Research Letters*, 20(6), 064029. <https://doi.org/10.1088/1748-9326/add6b7>
38. Steidinger, K. A. (2009). Historical perspective on *Karenia brevis* red tide research in the Gulf of Mexico. *Harmful Algae, Understanding the Causes and Impacts of the Florida Red Tide and Improving Management and Response*, 8(4), 549–561. <https://doi.org/10.1016/j.hal.2008.11.009>
39. Steidinger, K., Vargo, G., Tester, P., & Tomas, C. (1998). Bloom dynamics, and physiology of *Gymnodinium breve* with emphasis on the Gulf of Mexico. *Physiological Ecology of Harmful Algal Blooms*, 133–153.
40. Stumpf, R. P., Li, Y., Kirkpatrick, B., Litaker, R. W., Hubbard, K. A., Currier, R. D., Harrison, K. K., & Tomlinson, M. C. (2022). Quantifying *Karenia brevis* bloom severity and respiratory irritation impact along the shoreline of Southwest Florida. *Plos One*, 17(1), e0260755. <https://doi.org/10.1371/journal.pone.0260755>

41. Tomasko, D., Landau, L., Suau, S., Medina, M., & Hecker, J. (2024). An evaluation of the relationships between the duration of red tide (*Karenia brevis*) blooms and watershed nitrogen loads in southwest Florida (USA). *Florida Scientist* 87 (2).
42. USF Water Institute. (2026). *Welcome to the Water Atlas*. <https://wateratlas.org>
43. Vargo, G. A., Heil, C. A., Fanning, K. A., Dixon, L. K., Neely, M. B., Lester, K., Ault, D., Murasko, S., Havens, J., Walsh, J., & Bell, S. (2008). Nutrient availability in support of *Karenia brevis* blooms on the central West Florida Shelf: What keeps *Karenia* blooming? *Continental Shelf Research, Ecology and Oceanography of Harmful Algal Blooms in Florida*, 28(1), 73–98. <https://doi.org/10.1016/j.csr.2007.04.008>
44. Walsh, J. J., Jolliff, J. K., Darrow, B. P., Lenes, J. M., Milroy, S. P., Remsen, A., Dieterle, D. A., Carder, K. L., Chen, F. R., Vargo, G. A., Weisberg, R. H., Fanning, K. A., Muller-Karger, F. E., Shinn, E., Steidinger, K. A., Heil, C. A., Tomas, C. R., Prospero, J. S., Lee, T. N., ... Bontempi, P. S. (2006). Red tides in the Gulf of Mexico: Where, when, and why? *Journal of Geophysical Research: Oceans*, 111(C11). <https://doi.org/10.1029/2004JC002813>
45. Walsh, J. J., Weisberg, R. H., Lenes, J. M., Chen, F. R., Dieterle, D. A., Zheng, L., Carder, K. L., Vargo, G. A., Havens, J. A., Peebles, E., Hollander, D. J., He, R., Heil, C. A., Mahmoudi, B., & Landsberg, J. H. (2009). Isotopic evidence for dead fish maintenance of Florida red tides, with implications for coastal fisheries over both source regions of the West Florida shelf and within downstream waters of the South Atlantic Bight. *Progress in Oceanography*, 80(1), 51–73. <https://doi.org/10.1016/j.pocean.2008.12.005>
46. Wang, C., Wang, Z., Wang, P., & Zhang, S. (2016). Multiple Effects of Environmental Factors on Algal Growth and Nutrient Thresholds for Harmful Algal Blooms: Application of Response Surface Methodology. *Environmental Modeling & Assessment*, 21(2), 247–259. <https://doi.org/10.1007/s10666-015-9481-3>
47. Weisberg, R. H., Liu, Y., Lembke, C., Hu, C., Hubbard, K., & Garrett, M. (2019). The Coastal Ocean Circulation Influence on the 2018 West Florida Shelf *K. brevis* Red Tide Bloom. *Journal of Geophysical Research: Oceans*, 124(4), 2501–2512. <https://doi.org/10.1029/2018JC014887>
48. Weisberg, R. H., Zheng, L., & Liu, Y. (2016). West Florida shelf upwelling: Origins and pathways. *Journal of Geophysical Research: Oceans*, 121(8), 5672–5681. <https://doi.org/10.1002/2015JC011384>
49. Weisberg, R. H., Zheng, L., Liu, Y., Lembke, C., Lenes, J. M., & Walsh, J. J. (2014). Why no red tide was observed on the West Florida Continental Shelf in 2010. *Harmful Algae, Nutrient Dynamics of Karenia Brevis Red Tide Blooms in the Eastern Gulf of Mexico*, 38, 119–126. <https://doi.org/10.1016/j.hal.2014.04.010>
50. Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D. M., & Cochlan, W. P. (2015). Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae*, 49, 68–93. <https://doi.org/10.1016/j.hal.2015.07.009>
51. Yan, Z., Kamanmalek, S., & Alamdari, N. (2024). Predicting coastal harmful algal blooms using integrated data-driven analysis of environmental factors. *Science of The Total Environment*, 912, 169253. <https://doi.org/10.1016/j.scitotenv.2023.169253>
52. Yan, Z., Kamanmalek, S., Alamdari, N., & Nikoo, M. R. (2024). Comprehensive Insights into Harmful Algal Blooms: A Review of Chemical, Physical, Biological, and Climatological Influencers with Predictive Modeling Approaches. *Journal of Environmental Engineering*, 150(4), 03124002. <https://doi.org/10.1061/JOEEDU.EEENG-7549>

53. Zheng, X., Jia, G., Zhao, Y., & Yan, T. (2025). Involvement of four alga toxins in the risks of human neurodegenerative diseases: Toxicogenomic data mining and bioinformatics analysis. *Journal of Environmental Sciences*, 158, 151–164. <https://doi.org/10.1016/j.jes.2025.02.036>
54. Zohdi, E., & Abbaspour, M. (2019). Harmful algal blooms (red tide): A review of causes, impacts and approaches to monitoring and prediction. In *International Journal of Environmental Science and Technology* (Vol. 16). Center for Environmental and Energy Research and Studies. <https://doi.org/10.1007/s13762-018-2108-x>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.