REMapp Project Report

Carter Baker, Matthew Duus,

Hayden Parks, Callie Unterreiner

Advisor: Dr. Elshall

ABSTRACT

This study uses machine learning applications to analyze and predict *Karenia brevis* red tide blooms along the coast of Fort Myers, Florida, specifically in the Caloosahatchee River. Implementing add-ons into JupyterLab using Python, as well as sourcing data from USGS, NOAA, Water Atlas, and Copernicus Marine, allowed for in-depth analysis of the Caloosahatchee River's variable characteristics on an annual basis. Environmental data was classified as river discharge, precipitation, nutrient loading (consisting of total nitrogen and total phosphorus), *Karenia brevis* cell count concentration, wind speed and direction, and sea surface temperatures. The data were collected for the years 1993 to current (as of April 1st, 2025) and interpolated for values that were not measured due to external factors. The model initially showed perfect accuracy due to data leakage from HAB_Cells inclusion. After removing HAB_Cells and validating the data, the Random Forest model achieved 73% accuracy but struggled to detect blooms. Logistic Regression performed better at bloom detection (recall = 0.42) despite lower overall accuracy (59%). Results suggest that *Karenia brevis* blooms are linked to nutrient loading, discharge, wind, precipitation, and SST. However, improved forecasting will require broader, richer datasets to better capture environmental complexity and support bloom prevention efforts.

INTRODUCTION

Note: This study and its conclusions are based on the Caloosahatchee River in Fort Myers, Florida.

Annually, the Gulf Coast of Florida faces a harmful algal bloom event known as red tide. *Karenia brevis*, the culprit behind red tide in Florida, rapidly depletes oxygen and releases toxins into water systems. Results of this are mass fish and marine life kills due to oxygen dead zones and infection, contamination of shellfish causing them to be dangerous for consumption, respiratory irritation especially among those who are immunocompromised, and beach closures. The severity of blooms has been increasing in magnitude over the past 30 years and has presented a significant challenge for coastal living communities. Red tide degrades environmental quality, economic downturn, and decreases in quality of life. This study's primary goal was to develop a machine learning model that can predict blooms based on environmental conditions. Predicting blooms can benefit further research for prevention and solutions. Common contributors to red tide severity are river discharge, precipitation, nutrient loading, wind speed and direction, and sea surface temperatures.

METHODS

The initial foundation of this research stems from the local impacts of red tide to counties along the Gulf of America. The research implements Python and Machine Learning to demonstrate how red tide correlates to temporal variability within environmental conditions. The start of this research involved operating JupyterLab to run Python code for comprehensive analysis of data extracted from databases. Database sites such as USGS (United States Geological Survey), NOAA (National Oceanic and Atmospheric Administration), Water Atlas, HABSOS (Harmful Algal BloomS Observing System), and Copernicus Marine Data were used as data sources. As data was integrated into Python, additional packages were imported to support further analysis and construction of data. Packages included pandas, numpy, matplotlib.pyplot, matplotlib.dates, seaborn, geopandas, xarray, requests, os, shutil, and zipfile. Each subsection of data was constructed in separate files, which eventually was combined into a finalized file where machine learning analyzed correlations between the subsections. The subsections researched include Caloosahatchee River discharge, precipitation, nutrient loading (nitrogen and phosphorus), *Karenia brevis* cell count concentration, wind speed and direction, and lastly sea surface temperature.

Caloosahatchee River discharge data was obtained through USGS. The primary goal of obtaining this data was to gather mean daily values from 1993 to 2025, with measurements in terms of cubic feet per second. The USGS Station ID used was #2292900, which was located slightly inland within the Caloosahatchee River. A csv. file containing the necessary data was downloaded and then imported into Python. The 'Date' column was converted to datetime format and additional columns for 'Year' and 'Month' were added to facilitate seasonal and annual trend analysis. The data frame was constructed and used for graphical analysis. Missing discharge values presented as NaN were interpolated to maintain data integrity and prevent significant outliers from skewing the data. Annual peak discharge was isolated and used for correlation analysis, specifically linking peak discharges to hurricane season. This provided additional information on whether hurricanes are linked to annual peak discharge, which if correlated may play a factor in algal bloom severity.

Precipitation data for the Caloosahatchee River watershed was acquired from NOAA using the Global Historical Climatology Network Daily (GHCND) dataset. The primary station selected for this analysis was Fort Myers Page Field (Station ID: USW00012835), chosen for its geographic relevance and long-term data availability. Data was accessed through NOAA's Climate Data Online (CDO) API. Using Python's requests module, custom scripts were written to automate data retrieval for each year from 1993 through 2024, querying daily precipitation totals in millimeters. Precipitation values were processed using the pandas library converting raw JSON outputs to tabular form. The API response was flattened to table format separating fields for data, station ID, location, and precipitation depth. The data was cleaned from missing and negligible entries. Daily values were used for trend analysis and rolling averages were used to investigate rainfall event clustering. Using matplotlib and seaborn libraries daily precipitation time series were plotted to determine rainfall intensity and frequency, and cumulative annual rainfall plots were used to examine interannual variability and compare seasonal totals. These visualizations helped link precipitation variability to hydrological responses in the Caloosahatchee system and provided critical context for understanding nutrient transport, discharge rates, and potential triggers for red tide events.

Total nitrogen (TN) and total phosphorus (TP) data was sourced from Water Atlas, which supports water quality monitoring efforts within the Caloosahatchee River watershed. The data was in tabular format with timestamped entries and was imported into Python using the pandas library. Dates were converted to datetime format to facilitate time-series manipulation and correlation analysis with other environmental characteristics such as discharge, temperature, *K. brevis* concentration, etc. The TN and TP data was filtered to ensure missing values, NaN, were excluded and duplicate values were removed. The units for the concentration data were set to mg/L and monthly mean values for both TN and TP were calculated. This benefited long-term trend reviewal and was accomplished by extracting annual and monthly values. To visualize the data, matplotlib was used. This presented nutrient trends over time and created additional plots to compare concentrations of *K. brevis* and discharge to nutrient concentrations.

K. brevis cell count concentrations within the Caloosahatchee River and Gulf of America were obtained from the Harmful Algal Bloom Observing System managed by NOAA. This is a regional data system and focuses on water quality monitoring databases. The concentration of *K. brevis* was measured in cells/L and was recorded at various stations along the river and the Gulf. The data was imported into Python using pandas, and data construction such as datetime formatting changes were completed. Missing values were removed and concentrations near zero or below detection limits were filtered out, as the goal of this data was to review significant bloom events. Monthly averages were calculated and aligned with other environmental factors such as nutrient concentrations, meteorological conditions, and discharge. Line plots and time-series graphs were created which helped with identifying correlations between other variables. Data visualization presented potential lag time due to environmental conditions as well.

Wind data for the Caloosahatchee River region was obtained from NOAA using data sets provided by the integrated surface database (ISD). The selected station was the Fort Myers Page Field (Station ID: 722110-12835) due to its proximity to the S-79 structure in Olga and its long-term data availability (1993-present). Preprocessed data from NOAA ISD Lite was used, including datasets of daily averages of wind speed (m/s) and wind direction (degrees from North) based on hourly observations. Python packages including pandas, numpy, matplotlib, and requests were used in data preprocessing and analysis. Date fields were converted to datetime objects for time series analysis, missing or invalid data entries were removed, and the clean and structured data was saved and exported as a .csv file. Using matplotlib and seaaborn libraries the data was visualized with time series plots and histograms to observe long term trends and seasonal variability, additional plots like rolling averages and comparison overlays were added to make data more visual and interpretable. This workflow enabled structured, automated analysis of environmental wind data, facilitating integration with other parameters relevant to red tide prediction such as water temperature, discharge, and nutrient concentrations.

The sea surface temperature data used in this analysis was sourced from the Copernicus Marine Data Store. The data was provided in netCDF format, which is commonly used for storing multidimensional scientific variables. To retrieve the data a python script written by Copernicus was used, after which it was processed using the xarray library to structure it into an analyzable dataset. This dataset includes daily sea surface temperature measurements from depths ranging between 0.5 meters and 10 meters, covering a geographic extent from Charlotte Harbor down to the southern tip of Florida, including the Caloosahatchee River region. To analyze long-term patterns, the mean temperature was calculated across the latitude, longitude, and depth dimensions, generating a comprehensive view of spatially averaged sea surface temperature trends from 1991 to the present.

Machine learning models were trained and tested on all previously discussed environmental parameters, incorporating them into a fully integrated data set. After data structuring and formatting was completed, multiple processes were run to determine which model had the highest performance. Models included the Random Forest model and the Logistic Regression model, as well as a Correlation Matrix. Noticeable data leakage occurred initially, skewing the results to appear much more performative. Once environmental parameters were refined, both models exhibited declines in performance, but presented much more realistic results. Pandas, numpy, and scikit-learn libraries were used within the machine learning process of red tide bloom occurrence research.

RESULTS AND DISCUSSION

SECTION 1

Line Plot Data Visualization Per Environmental Characteristics [Figure's 1 – 10]

Figure 1: Stacked Subplots Containing Environmental Parameters

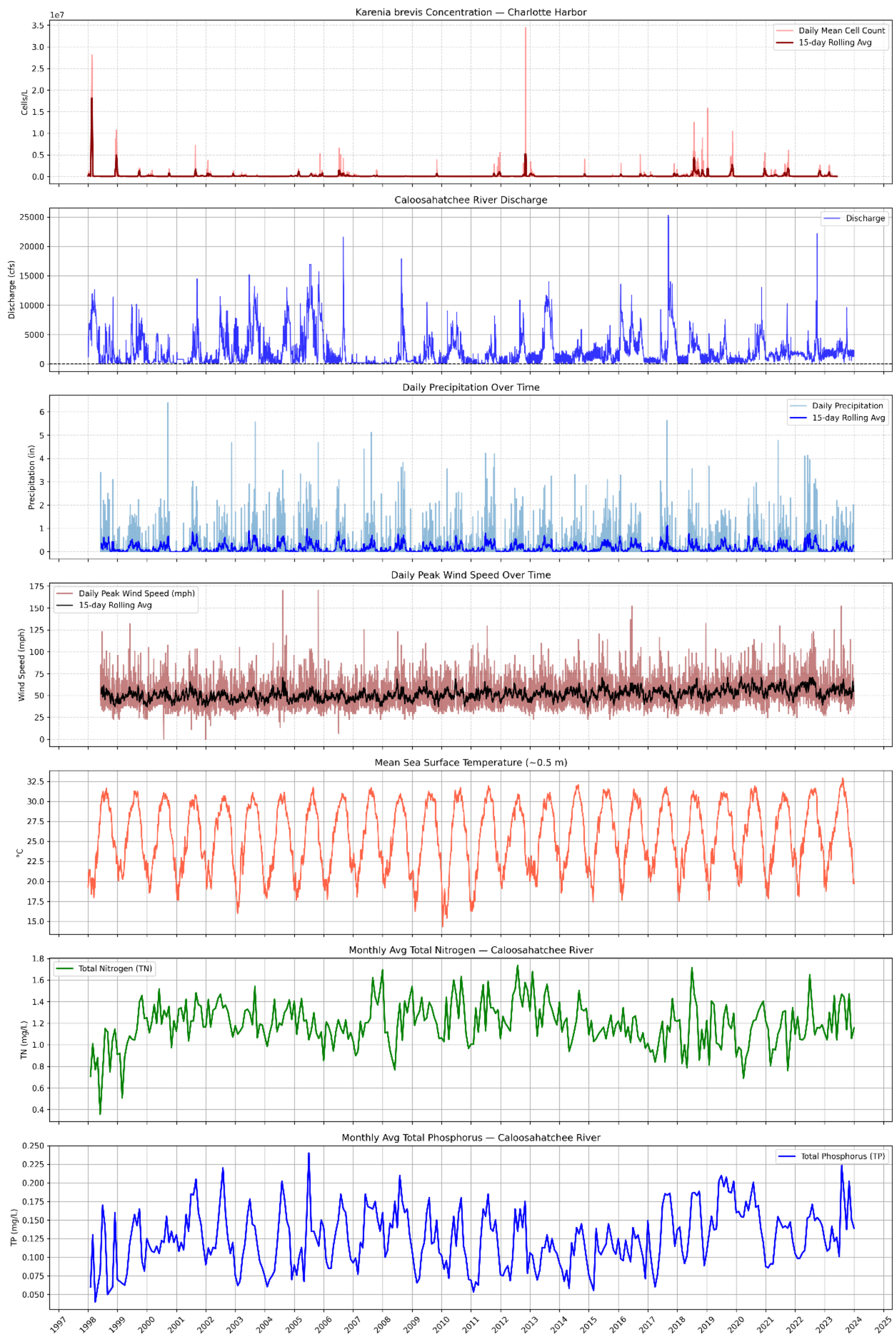


Figure 1 demonstrates environmental parameter subplots stacked for visualization purposes. The contribution of this data was significant for building a base for machine learning to predict bloom occurrences. The plot extends from 1998 to 2025, providing data for each environmental parameter. Subplot 1 provides insight into bloom events, with notable bloom events being characterized by a sharp spike in cell count. This appears to be correlated with other parameters such as spikes in discharge and precipitation.

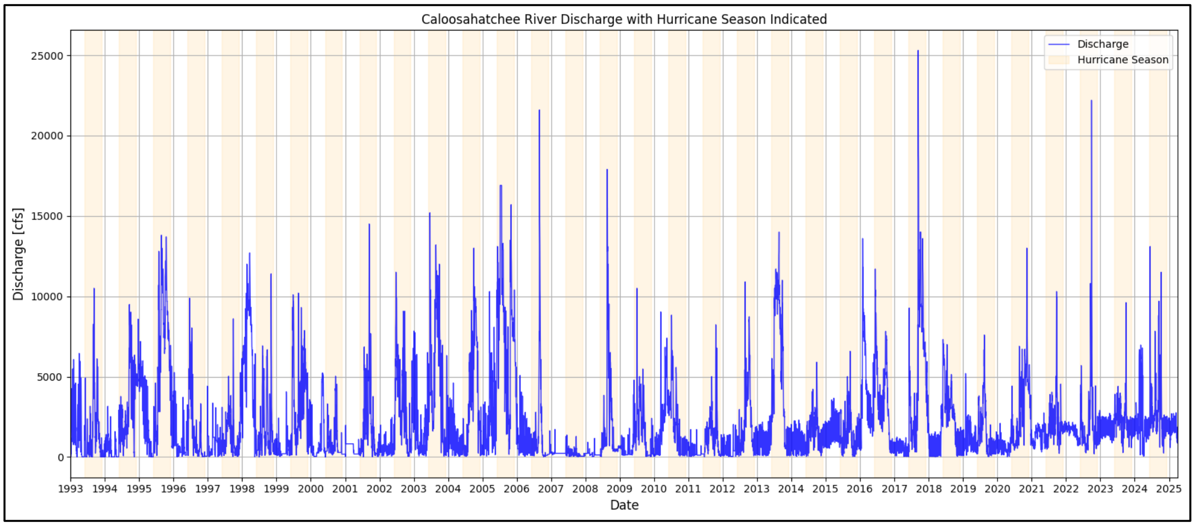
Discharge is a significant parameter when considering the Caloosahatchee River and its surrounding red tide blooms. Discharge is the primary transport source for nutrient loading to occur in the Gulf, which can result in eutrophication. Eutrophication is nutrient loaded waters resulting in algal blooms. Precipitation is heavily correlated with discharge. When rainfall intensity increases, discharge also increases due to a large volume of runoff draining into the Caloosahatchee River. All areas within reasonable distance of the Caloosahatchee River drain into it because it is the main source of water discharge within the watershed. Increased rainfall carries nutrients within the runoff, which end up in the Gulf and promote bloom intensity and occurrences.

Subplot 4 is the wind parameter. Winds tend to fluctuate based on weather conditions and atmospheric conditions. Wind affects coastal upswelling and water column mixing, which can lead to bloom events persisting over long periods of time and possibly being transported along the coast. Subplot 5 demonstrates sea surface temperatures, which appear to remain relatively consistent and are heavily correlated to seasonal patterns. Bloom events generally occur during the summer, but last well into the fall and possibly winter. With sea surface temperatures remaining elevated during those seasons, blooms are more likely to appear and persist, as *Karenia brevis* are known to prefer warmer waters.

The most significant parameter is nutrient concentration, which is visible in Subplot 6 and 7. Monthly averages of both TN and TP were accounted for from 1998 to 2025. This data was necessary as nutrient loading is one of the primary factors in bloom occurrences, severity, and persistence. As nutrient concentrations peak, it is notable that blooms appear to occur, leading to a strong likelihood of correlation between the two.

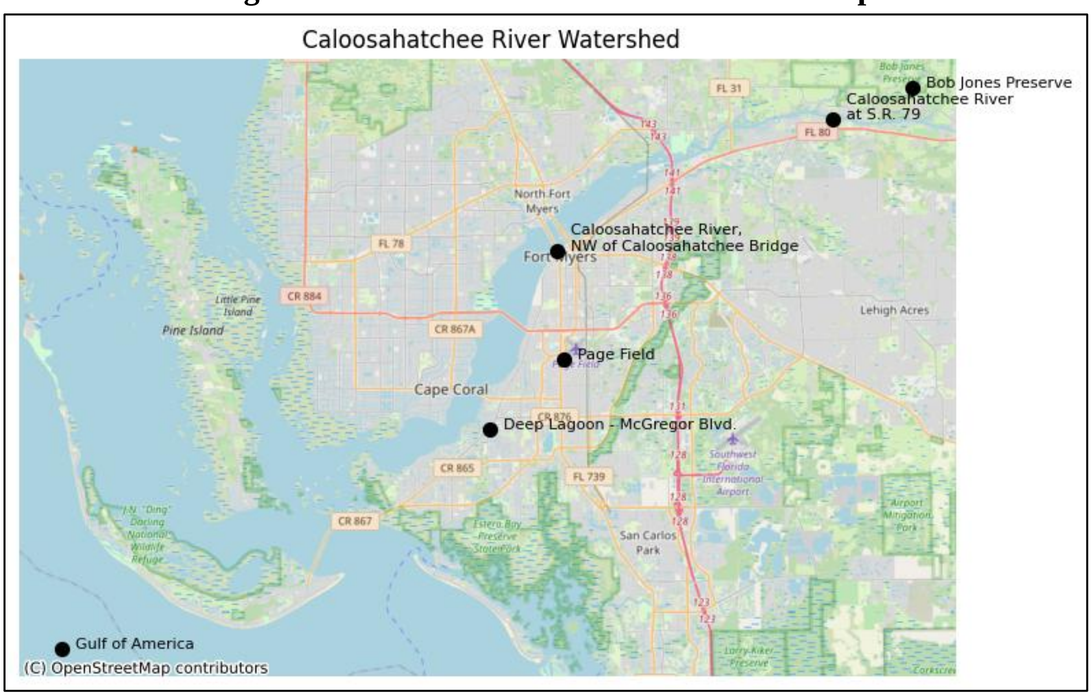
Environmental parameters are further reviewed within the machine learning portion of the research and correlation significance is measured, as well as bloom prediction accuracy and precision.

Figure 2: Caloosahatchee River Discharge with Hurricane Season Indicated Line Plot



Discharge significance plays a major role in the Caloosahatchee River's hydrology characteristics. *Figure 1: Subplot 2* and *Figure 2* line plots demonstrate major peaks each year, which have been linked to hurricane season. This claim is based on Python code that accounted for peak discharge annually and the time span hurricane season occurs. The code portrayed an 81.25% chance of peak discharge occurring during hurricane season. If correlation significance is noticeable between discharge and algal bloom severity, then hurricane season could be a major factor in determining whether a severe bloom will occur during the season.

Figure 3: Caloosahatchee River Watershed Map



Visualization of selected source monitoring sites was necessary for understanding the effects of nutrient migration inland to the Gulf. *Figure 3* was developed using geopandas along with defined site coordinates. This presented a visual of the watershed under review and provided a visual location for each source monitoring site.

Figure 4 through Figure 8 were designed using CartoPy and matplotlib:

Figure 4: Multi-Year Average Sea Surface Temperature (0.5m depth) Multi Color Map

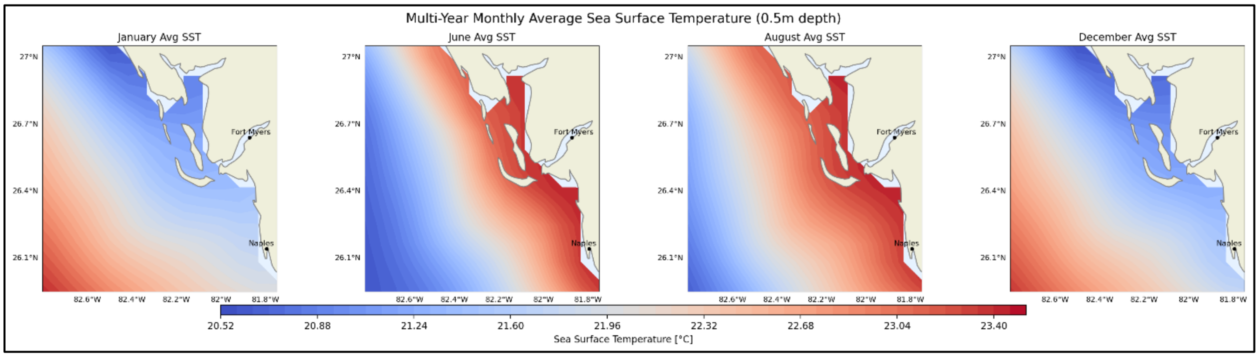
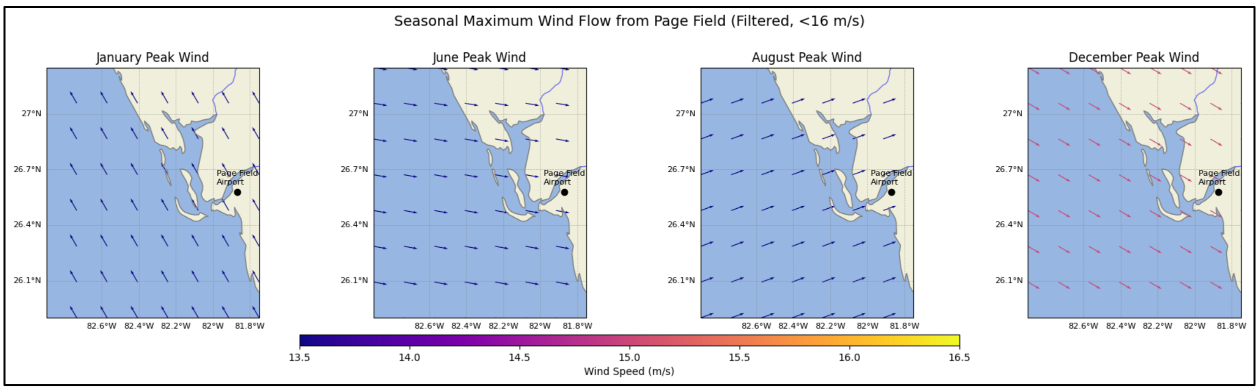


Figure 4 represents the average sea surface temperature from 1998 to 2025. Sea surface temperatures are heavily influenced by seasons. This is notable in the winter months where the ocean is blue toward the coast indicating cool temperatures and in the summer months where the ocean is red indicating warm temperatures. The warm waters within the summer months align with seasonal blooms, which could promote red tide initiation and intensification. Providing sea surface temperatures will allow for machine learning to develop correlation between blooms and water temperatures. This gives an idea of what temperature ranges *K. brevis* may thrive in.

Figure 5: Seasonal Maximum Wind Flow from Page Field (Filtered, <16 m/s) Multi Vector Plot



Wind data in *Figure 5* shows seasonal wind direction and speed. The wind speed appears to remain relatively constant with the outlier being December. Wind speeds rise approximately 1.5 m/s in December which could be due to wind shear and jet stream conditions varying. Throughout the summer months, the wind direction is primarily East. This is significant because it promotes onshore transport of nutrients, which leads to coastal blooms occurring. Possibilities of enhanced blooms due to easterly wind is significantly higher when nutrients are being transported towards the coast and held there. The seasonal patterns suggest that wind speed and direction are necessary for the machine learning model to help account for physical mechanisms influencing bloom movement and severity along the coast.

Figure 6: Seasonal *Karenia brevis* Blooms Multi Dot Plot

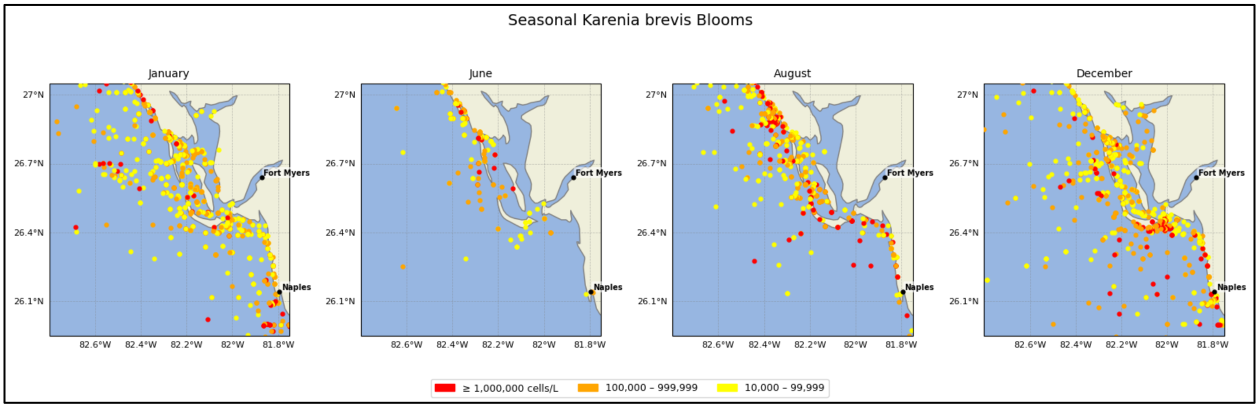
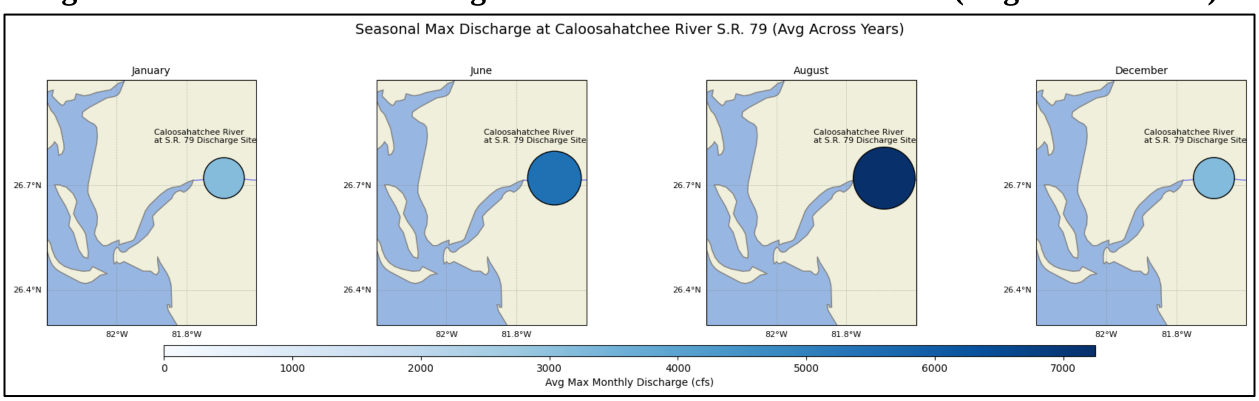


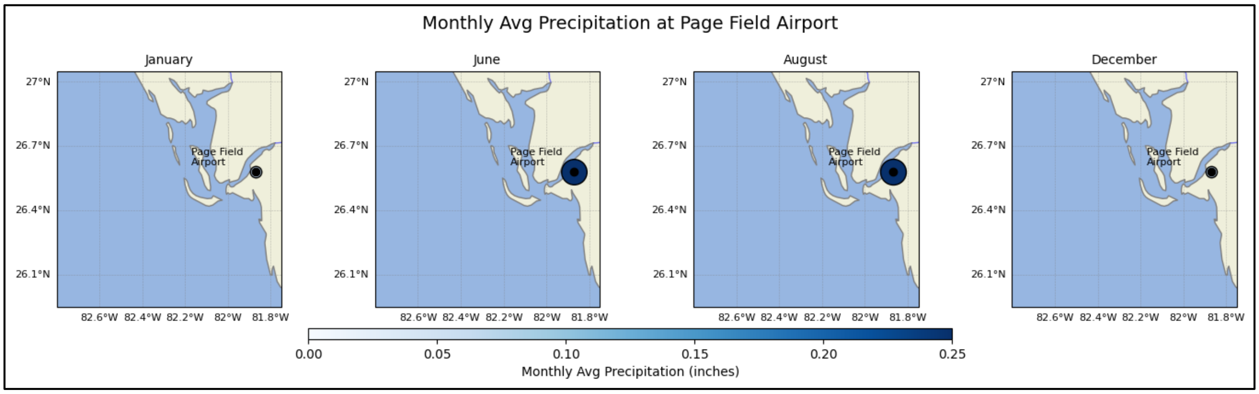
Figure 6 presents every bloom event that has occurred between 1953 to 2025. The data was separated into seasonal graphics, which gave insight into which months experience the most intense blooms. Based on the graphics, June appears to be the start of red tide blooms, which further intensify throughout the summer until August, where peak concentrations are reached. Blooms are persistent throughout the months leading to the end of December, when concentrations start to decline until early summer. The persistence of blooms throughout nearly half of the year is likely influenced by environmental parameters such as nutrient inputs, hydrodynamics, and sea surface temperature. Coastal clustering is apparent in the August subplot, where the northern part of the graphic contains multiple bloom events with over 1,000,000 cells/L present. This data is valuable for machine learning's prediction and accuracy levels as it provides information that will allow it to conclude whether a bloom event is present or not.

Figure 7: Seasonal Max Discharge at Caloosahatchee River S.R. 79 (Avg Across Years)



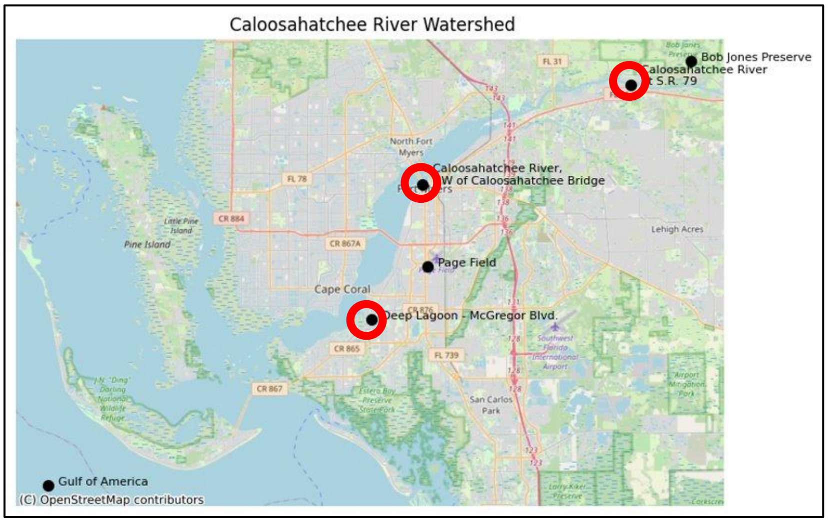
Reviewing seasonal max discharge within the Caloosahatchee River is necessary as it provides information on increases in nutrient loaded deposits into the Gulf. As seen in *Figure 7*, max discharge varies throughout the year, with its peak occurring at the peak of hurricane season. During wet months, where discharge and precipitation are elevated, there is increased transport in nutrients from inland Florida to the Gulf. The Caloosahatchee River's hydrological role is significant for the watershed, as it directly relates to the severity of blooms.

Figure 8: Monthly Avg Precipitation at Page Field Airport



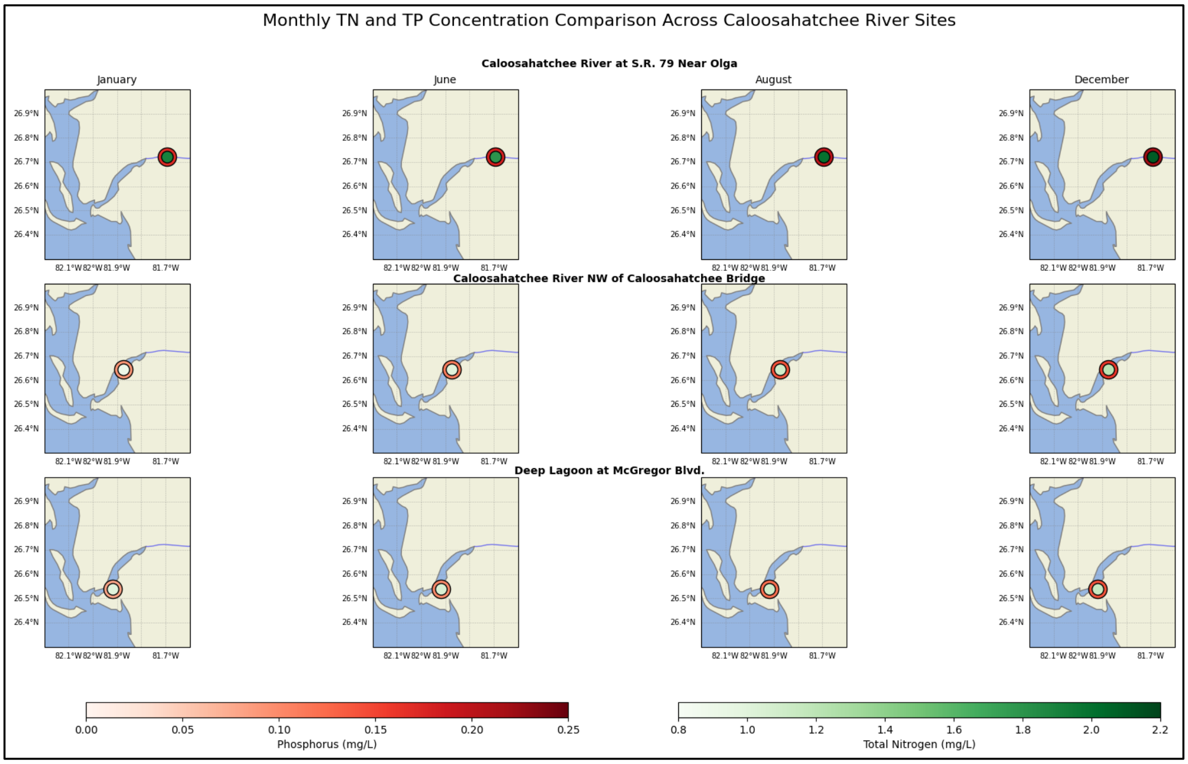
Understanding seasonal rainfall patterns is necessary for determining precipitation contributions to red tide. The start of wet season begins in June, which is noticeable in the June graphic, where an increase in precipitation is present. Based on the four graphics in *Figure 8*, August shows the highest average precipitation. This corresponds to Florida's wet season and also happens to fall within the peak of hurricane season, which can be linked to peak discharge occurring in the Caloosahatchee River. As expected, January and December demonstrate low precipitation due to dry season. Precipitation is significant because it is directly related to runoff. Runoff is the main source of nutrient transport and loading in rivers. The Caloosahatchee River discharges into the Gulf of America, so as precipitation increases, the nutrient concentrations also increase. The use of Page Field Airport data ensures that there is local data collected and provides valuable information for correlation analysis within machine learning.

Figure 9: TN and TP Monitoring Stations – Caloosahatchee



Exploration of TN and TP concentrations throughout the Caloosahatchee River was conducted in three sites: S.R. 79 Near Olga, Northwest Caloosahatchee Bridge, and Deep Lagoon. These sites provide data ranging from inland to the Gulf of America, which ensures a wide selection of data throughout the river.

Figure 10: Monthly TN and TP Concentration Comparison Across Caloosahatchee River Sites



When reviewing monthly total nitrogen and total phosphorus concentration across Caloosahatchee River Sites, it is notable that inland water contamination is much more concentrated compared to the Gulf discharge of the river. S.R. 79 Near Olga appears to experience significantly higher concentrations of both TN and TP compared to the other sites. Many factors could contribute to this, but the most likely factor would be runoff from residential, commercial, and industrial sites eventually draining into the river. The runoff carries contaminants to the river such as pesticides, car oils, and particulates, which increase the chances of eutrophication inland. As the river flows toward the Gulf of America, the concentrations of TN and TP decrease, likely due to dispersion and dilution; however, this still fuels the Gulf with the necessary ingredients for red tide – warm water and nutrient loading. It is also important to note that the total nitrogen appears to be consistently increasing throughout the year at both Caloosahatchee Bridge and Deep Lagoon sites.

SECTION 2

Machine Learning Development and Operation:

Figure 11: Random Forest

```
Confusion Matrix:
[[38  3]
 [22  0]]

Classification Report:
              precision    recall  f1-score   support

           0       0.63      0.93      0.75        41
           1       0.00      0.00      0.00        22

    accuracy                           0.60        63
   macro avg       0.32      0.46      0.38        63
weighted avg       0.41      0.60      0.49        63
```

Figure 11 shows the Random Forest classifier resulted in strong performance identifying non-bloom conditions but failed entirely to classify any bloom events correctly. The confusion matrix revealed that 38 non-bloom months were correctly identified, while only 3 were misclassified as blooms. However, all 22 actual blooms were misclassified as non-blooms, resulting in zero true positives. As a result, the classification report shows a recall of 0.93 and a precision of 0.63 for the non-bloom class, but a precision, recall, and F1-score of 0.00 for the bloom class. The overall accuracy was 60%, which is misleadingly high due to the model's bias toward predicting the majority class. These results reinforce the issue of class imbalance and suggest that the model, as currently implemented, is unable to learn patterns associated with bloom conditions based on the selected environmental variables. Therefore, we trained a second model *(Figure 12)* to evaluate the robustness of the Random Forest classifier and gain further insight into the predictability of bloom events.

Figure 12: Logistic Regression

```
Accuracy: 0.60
Confusion Matrix:
[[26 15]
 [10 12]]

Classification Report:
              precision    recall  f1-score   support

           0       0.72      0.63      0.68        41
           1       0.44      0.55      0.49        22

    accuracy                           0.60        63
   macro avg       0.58      0.59      0.58        63
weighted avg       0.63      0.60      0.61        63
```

Unlike the Random Forest model, which failed to identify any bloom conditions, the Logistic Regression model—trained with class weighting to address data imbalance—achieved a more balanced performance. It correctly identified 12 out of 22 blooms and 26 out of 41 non-blooms, resulting in an overall accuracy of 60%. Although the precision for bloom predictions was moderate (0.44), the model achieved a recall of 0.55, indicating it was able to detect more than half of the actual bloom events. These results suggest that Logistic Regression is better suited to the class imbalance present in the dataset and can uncover meaningful relationships between environmental variables and bloom risk. While the model is not yet highly accurate, its ability to detect both classes makes it a stronger baseline for future model development and refinement.

To better assess the performance of our improved Logistic Regression model, we used ROC-AUC and precision-recall curves in addition to standard metrics like the confusion matrix and accuracy. These visual tools offer a deeper look at model behavior, particularly in imbalanced datasets like ours, where bloom events (1s) are much less frequent than non-blooms (0s). This step helped clarify whether the model is effectively capturing bloom dynamics or if it is still biased toward the majority class, despite improved recall for bloom events observed in earlier evaluations.

The resulting metrics indicated perfect performance: both the ROC curve and PR curve showed an area under the curve (AUC) of 1.00, and the classification report reflected 100% accuracy, precision, and recall across both bloom and non-bloom classes. While these results appear ideal, such flawless classification is highly improbable in real-world environmental datasets. This strongly suggests a problem such as data leakage, duplicate entries across training and testing sets, or another structural issue in the model pipeline. In practice, HAB prediction is influenced by a complex set of environmental variables with natural variability and uncertainty. Thus, although the visual outputs appear impressive, they must be interpreted with skepticism and prompt a careful review of the model setup and data handling to ensure validity.

We targeted the data leakage using the Correlation Matrix (*Figure 13*), revealed a strong positive correlation ($r = 0.61$) between HAB_Cells and the target variable HAB_Bloom, confirming that HAB_Cells directly encodes the bloom classification. Since HAB_Bloom was derived by thresholding HAB_Cells, including both in the model introduces label leakage, allowing the model to “cheat” by learning the outcome from its definition rather than from independent environmental indicators. This explains the unrealistically perfect model performance observed earlier. To ensure valid predictions based solely on external drivers like nutrient levels, SST, and discharge, we removed HAB_Cells from the feature set and re-ran the model.

Figure 13: Correlation Matrix

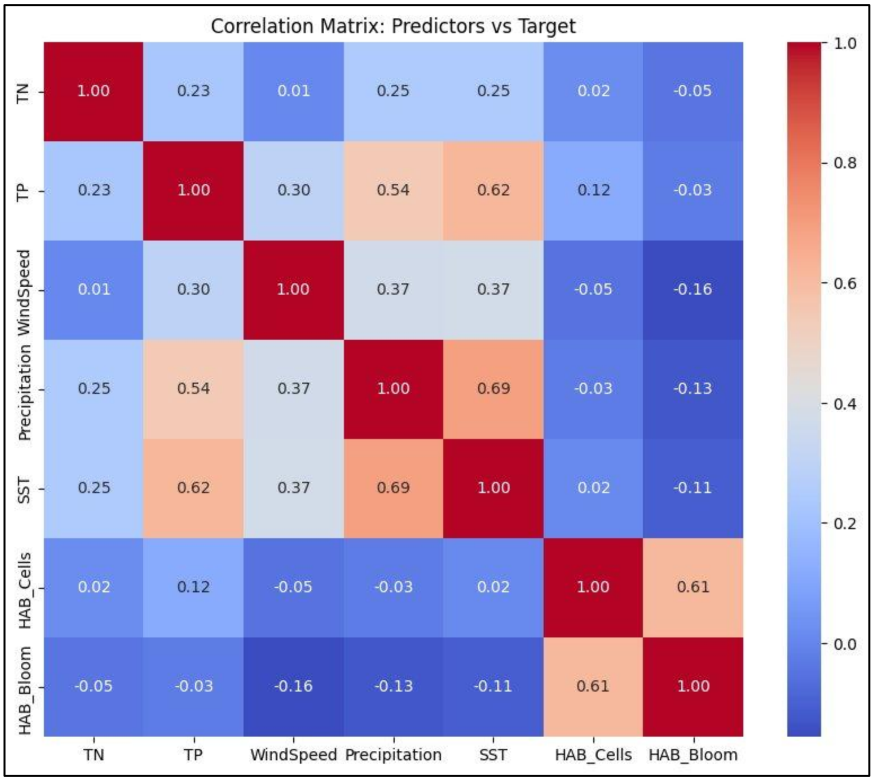
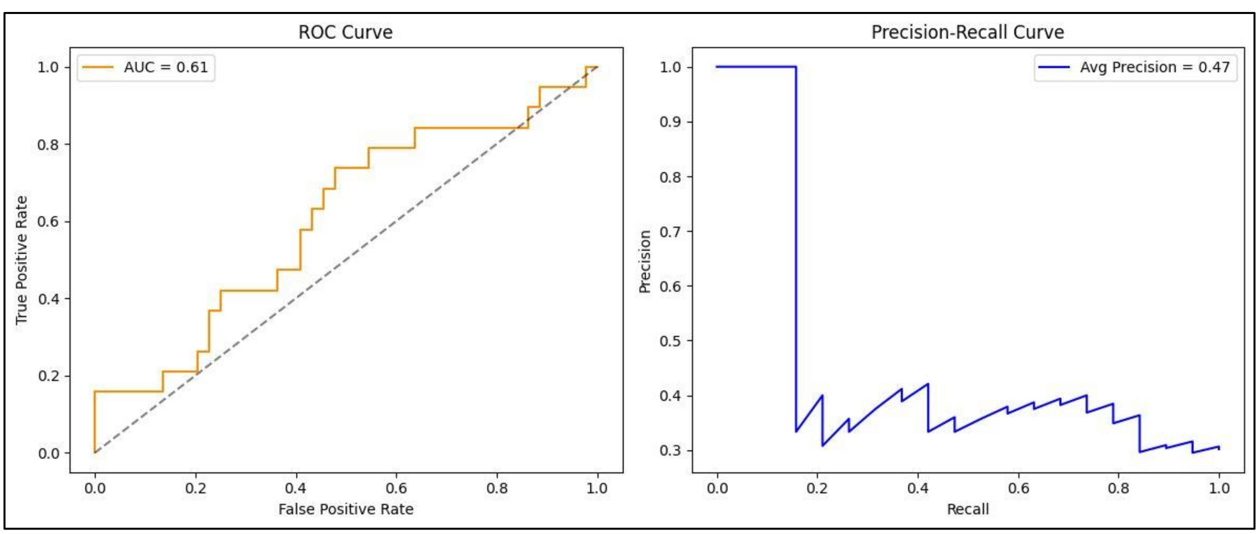


Figure 14 shows more realistic results for the ROC-AUC and precision-recall curves. With an ROC AUC of 0.61 and an average precision of 0.47, the model performs only slightly better than random and struggles to maintain precision as recall increases. These metrics align with earlier classification results and reinforce the challenge of predicting HABs using limited environmental variables.

Figure 14: Model Accuracy and Precision Curves



After removing HAB_Cells from the feature set to eliminate label leakage, we retrained the Random Forest classifier using only independent environmental variables. The model's performance increased from 60% to 73%. While it maintained strong precision and recall for non-bloom conditions, it struggled to correctly identify bloom events, detecting only 3 out of 19 (recall = 0.16).

This result reflects the inherent difficulty of predicting rare and complex phenomena like HABs using environmental drivers alone. Importantly, the removal of HAB_Cells restored the integrity of the modeling process and provided a more accurate baseline for evaluating model effectiveness in real-world conditions.

Figure 15: Random Forest Matrix (Post Data Leakage Fix)

```
Confusion Matrix:
[[43  1]
 [16  3]]

Classification Report:
              precision    recall  f1-score   support

           0       0.73      0.98      0.83        44
           1       0.75      0.16      0.26        19

    accuracy                           0.73        63
   macro avg       0.74      0.57      0.55        63
weighted avg       0.74      0.73      0.66        63
```

CONCLUSIONS

This project explored the relationship between environmental drivers and *Karenia brevis* red tide blooms along the Caloosahatchee River using machine learning models trained on over three decades of environmental data. Key findings reveal that while variables such as total nitrogen, total phosphorus, discharge, wind speed, precipitation, and sea surface temperature exhibit some predictive value, the current dataset lacks the resolution and diversity required for high-confidence bloom forecasting. Notably, the Random Forest model achieved 73% accuracy overall but struggled to identify bloom events, whereas Logistic Regression showed better sensitivity to blooms (recall = 0.42), highlighting the trade-off between overall accuracy and bloom detection.

The main contributions of this study include the integration of large-scale environmental datasets from multiple federal and regional agencies, the automation of data processing workflows in Python, and the implementation of machine learning within a JupyterLab environment to analyze and visualize red tide-related patterns. These efforts collectively demonstrate a scalable framework for red tide prediction that can be expanded and refined as more data becomes available.

Several limitations were encountered throughout the research, including an initial issue with data leakage due to the inclusion of HAB_Cells, which inflated early model performance. Additionally, the complexity of combining diverse environmental parameters into a coherent machine learning pipeline presented technical and conceptual challenges. A further limitation involved the steep learning curve associated with implementing and interpreting machine learning models.

Looking ahead, future research should prioritize expanding the dataset both temporally and spatially, and incorporating additional environmental indicators such as Gulf salinity, sea surface height, and the influence of the Loop Current. Incorporating higher-frequency data and biological indicators could significantly enhance model accuracy and reliability. These steps will be essential for developing a predictive tool that is not only accurate but actionable for environmental management and bloom prevention in the Gulf of America.

DATA REPOSITORY

The project's full codebase is available on GitHub at https://github.com/mkduus/caloosahatchee-red-tide-analysis. This repository contains the Jupyter Notebooks used for data cleaning, mapping, and machine learning analysis; the data files; the final presentation; and a copy of this final report.

REFERENCES

CMEMS. Copernicus Marine Service. (2025, April 16). https://marine.copernicus.eu/

Florida Red Tide Mitigation & Technology Development initiative. Mote Marine Laboratory & Aquarium. (2024, May 29). https://mote.org/research/centers-of-excellence/red-tide-initiative/

NOAA. (n.d.). National Oceanic and Atmospheric Administration. https://www.noaa.gov/

Red Tide. Florida Fish and Wildlife Conservation Commission. (n.d.). https://myfwc.com/research/redtide/statewide/

Protecting Florida Together. Red Tide | Water Quality. (n.d.). https://protectingfloridatogether.gov/resources/red-tide

USF Water Institute. (n.d.). *Water Atlas*. wateratlas.org. https://wateratlas.usf.edu/

USGS. (2025, April 19). https://www.usgs.gov/