Forecasting Algal Blooms

Erdos Institute Data Science Bootcamp 2024

Deniz Olgu Devecioglu Peter Doze Nathanial Lowry

(Mentor : Elizabeth Campolongo)

THE ERDŐS INSTITUTE

Helping PhDs get and create jobs they love at every stage of their career.

Introduction and Problem Statement

A harmful algal bloom (HAB) occurs when toxin-producing algae grow excessively in a body of water. HABs have serious economic impacts on local fisheries and are associated with harmful health effects in humans.

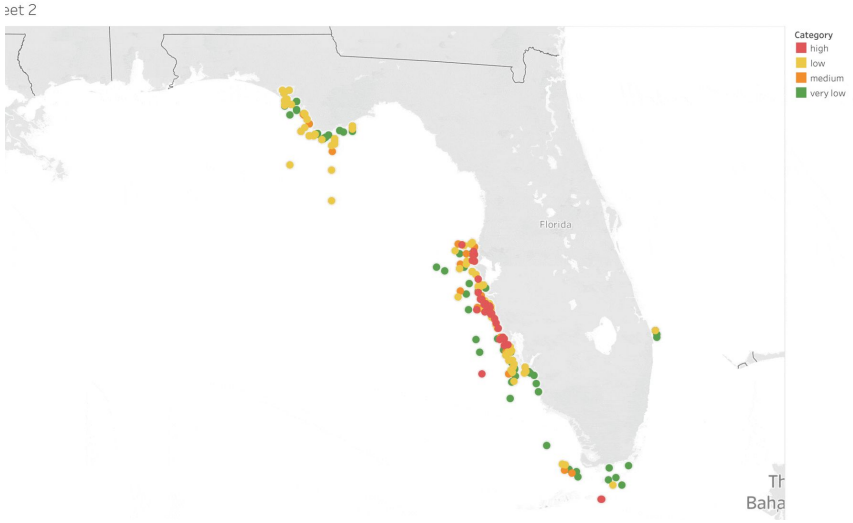
We want to know:

- What ecological features predict HABs?
- Can we use the available data to predict future HABs?

Data Collection

The National Oceanic and Atmospheric Administration (NOAA) has been sampling *K. Brevis* algae in the Gulf of Mexico since the early 1950s. The data they collect includes:

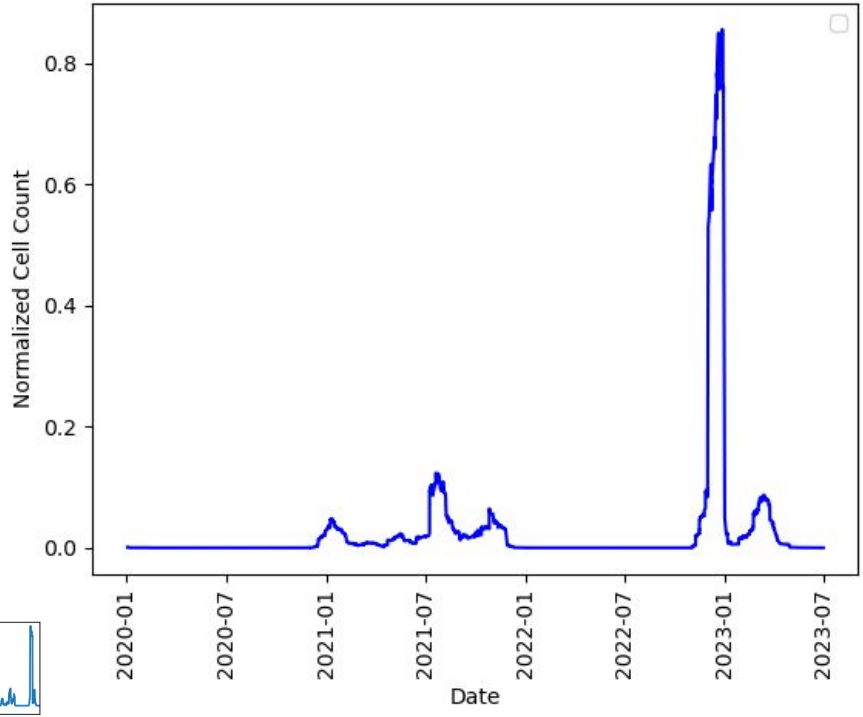
- Date and location of sample
- Cell count
- Water temperature
- Water salinity
- Wind speed/direction

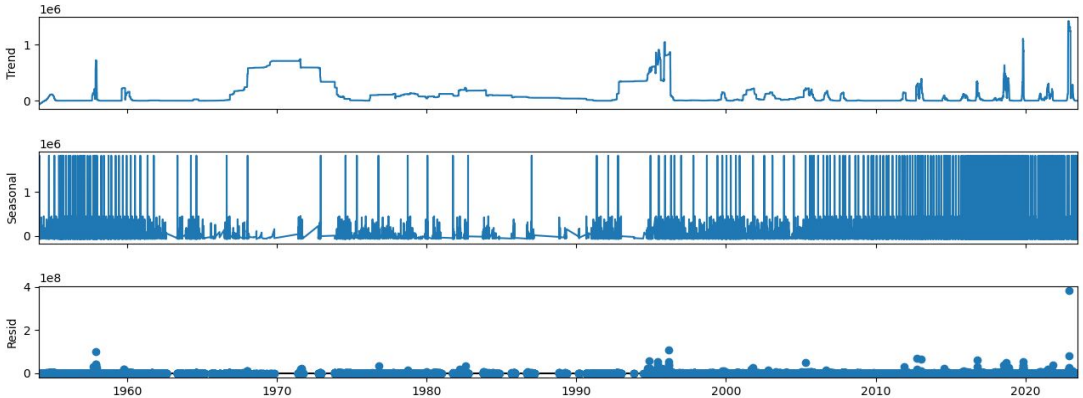


Exploratory Data Analysis (cont.)

Normalization: X - min / (max - min)

ADF and KPSS test



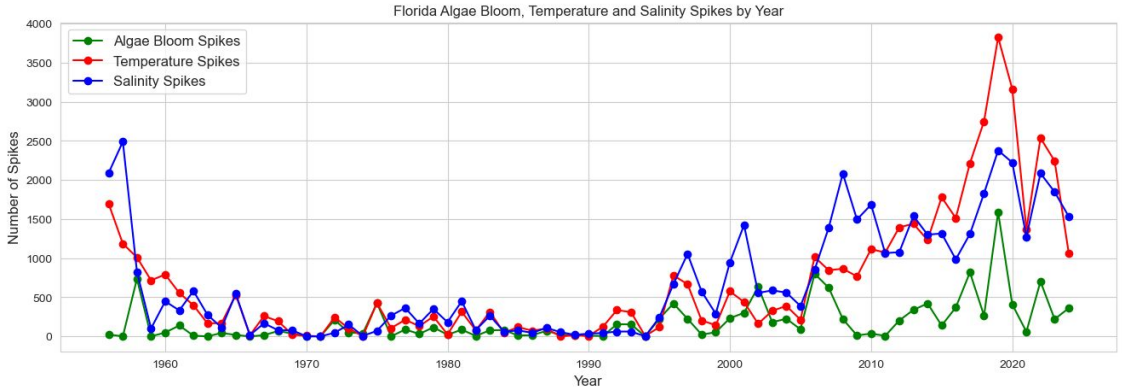


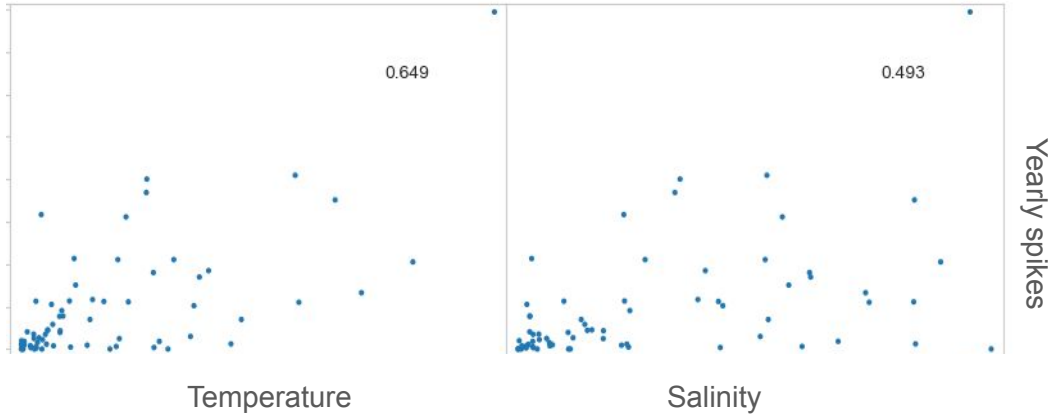
Exploratory Data Analysis

Predictive features

- Water temperature
- salinity

Both appear to be correlated to the number of yearly “spikes”, where a spike is defined to be any measurement that exceeds the average cell count up to the date of collection





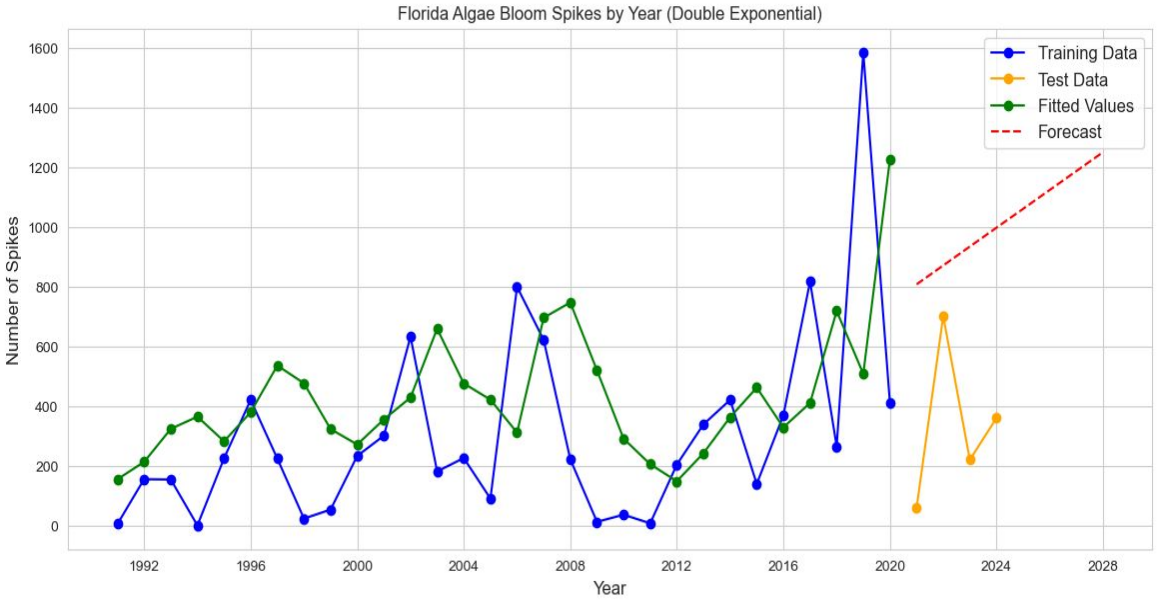
Results/Forecast: Number of Spikes

cv RMSE of the Baselines

	Base Model	RMSE
0	Average Base	360.487543
1	Naive Base	354.065375
2	Trend Base	340.714471
3	Random Walk with Drift	351.795166

cv RMSE of the Models

	Model	cv RMSE
0	Rolling Average	335.194830
1	Double Exponential Smoothing	320.443000
2	ARIMA	343.995000
3	ARIMA (AIC)	352.813458
4	Random Forest (Temp)	354.958910
5	Random Forest (Sal)	387.795886



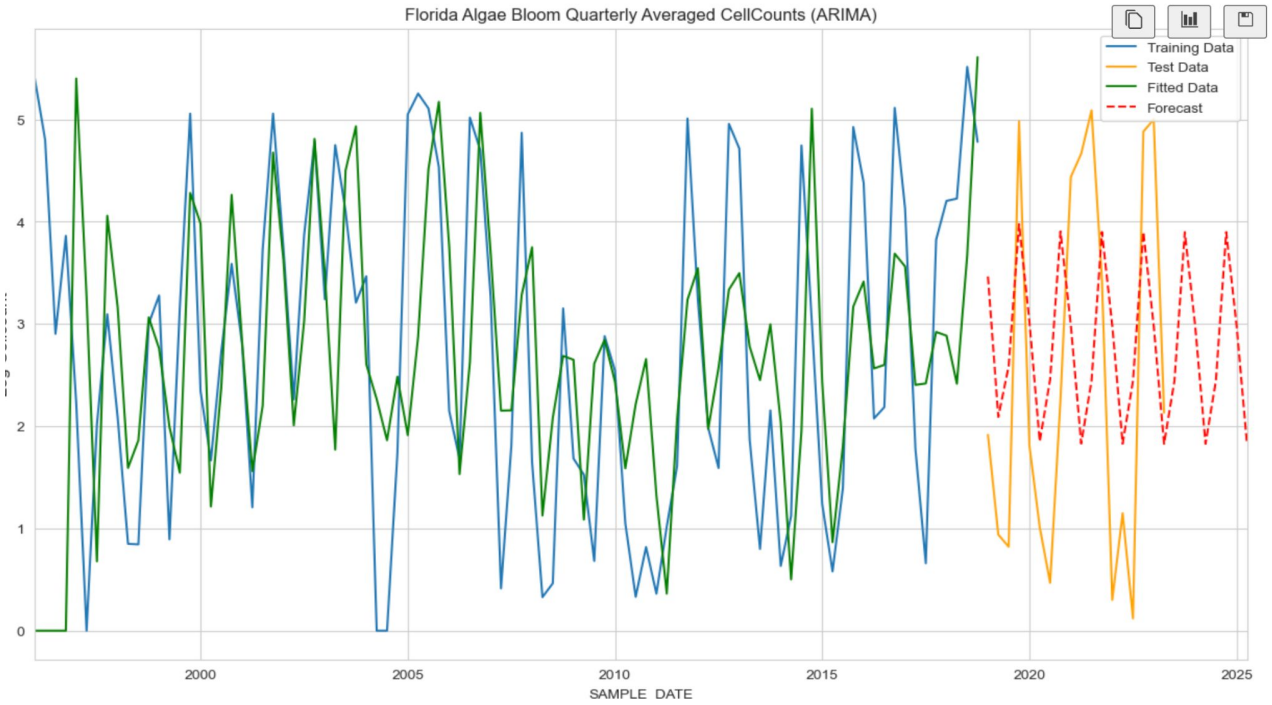
Results/Forecast: Algal Bloom (Quarterly)

cv RMSE of the Baselines

	Base Model	RMSE
0	Average Base	1.628660
1	Sesonal Base	1.355350
2	Naive Base	1.965576
3	Seasonal Naive	1.870055

cv RMSE of the Models

	Model	cv RMSE
0	Rolling Average	1.668450
1	Exponential Smoothing	1.620100
2	ARIMA	1.206855
3	Random Forest (Temp)	1.748257



Next Steps

- More effective use of Random Forests with the water temperature as feature.
- The NOAA documentation suggests that wind plays a factor in spreading algae from FL.
- Algal blooms in MS and TX.
- More consistent wind data needed.

Thank you!