Operational Forecasting of Harmful Algal Blooms in Florida Lakes Using a Two-Stage Bayesian Model

April 8, 2026

Jonathan Holt
Natalie Reynolds

RTI INTERNATIONAL

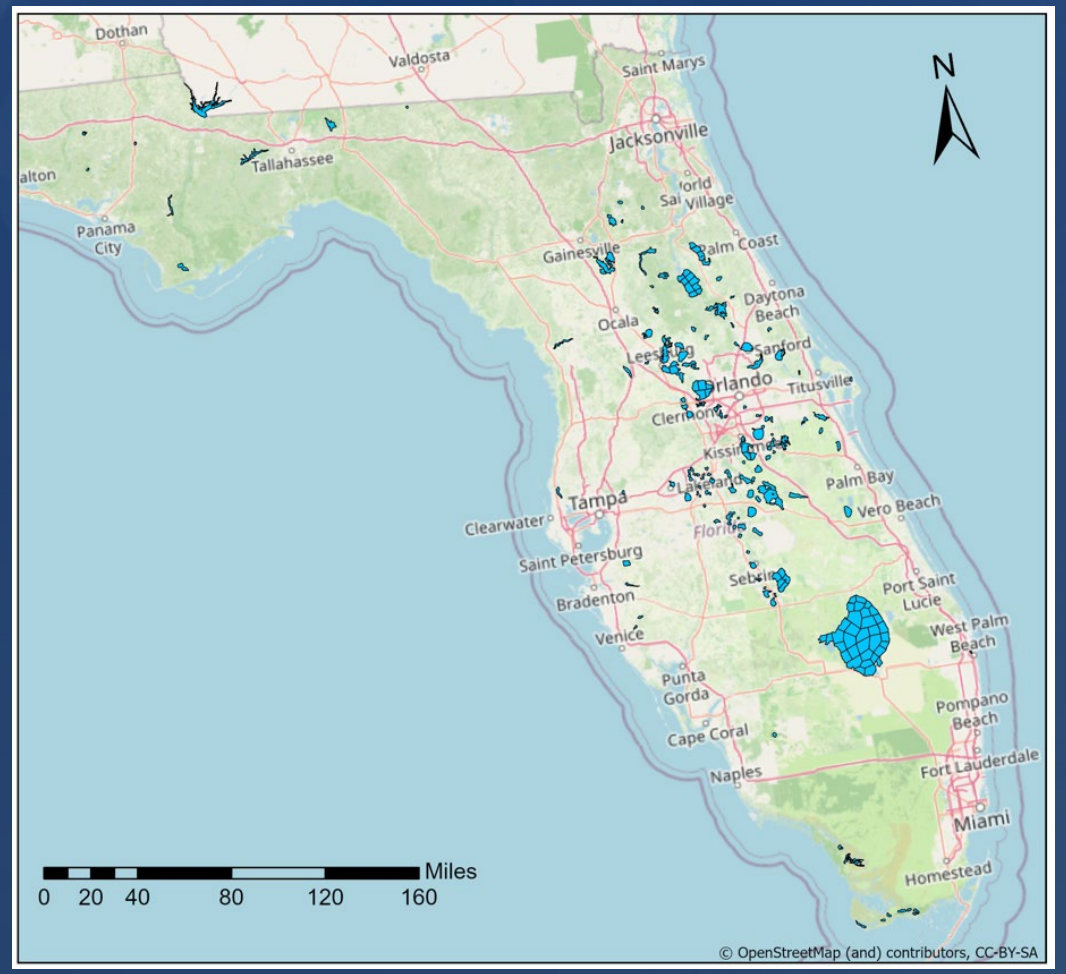
Motivation – Defining the Problem

- **Current Process**: The DEP responds to Harmful Algal Blooms (HABs) by collecting samples after they occur to inform the public.
- **The Problem**: The DEP wants to shift to a more *proactive* strategy for managing HABs.
- **Project Goal**: To predict when and where a HAB is likely to occur in the future, allowing DEP to use this information to preemptively monitor and/or mitigate HABs.

Primary modeling objective: Forecast 7-day HABs Magnitude

- Spatial extent: Satellite resolvable lakes in Florida
  - 134 total lakes
  - 8 of the largest lakes were divided into sub-lake zones:

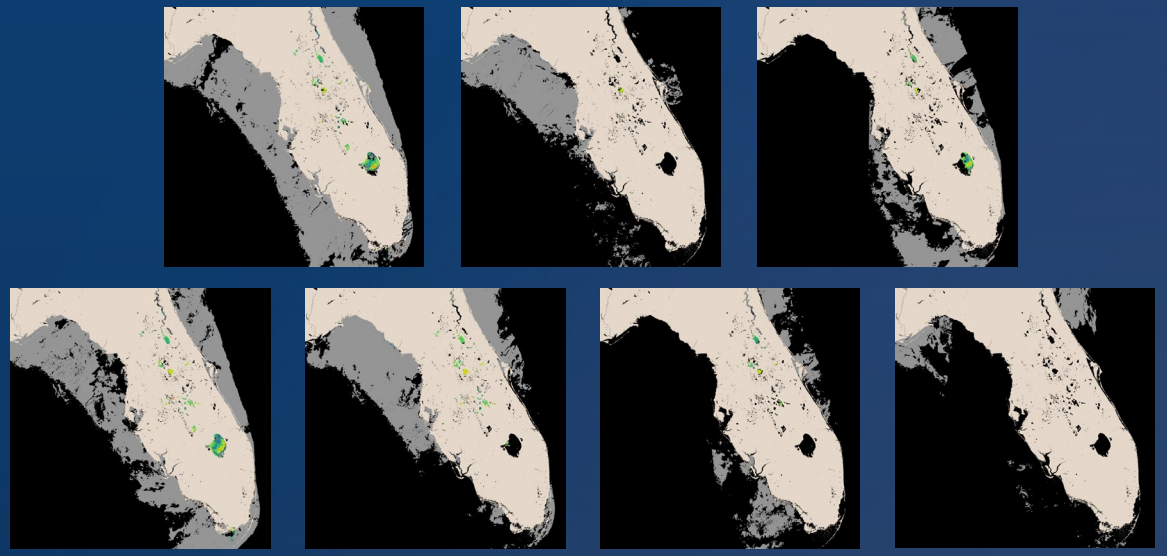
Lake	Number of Zones
Lake Okeechobee	34
Lake George	9
Lake Apopka	5
Lake Istokpoga	5
Crescent Lake	4
Lake Kissimmee	4
Lake Harris	3
Lake Tohopekaliga	3



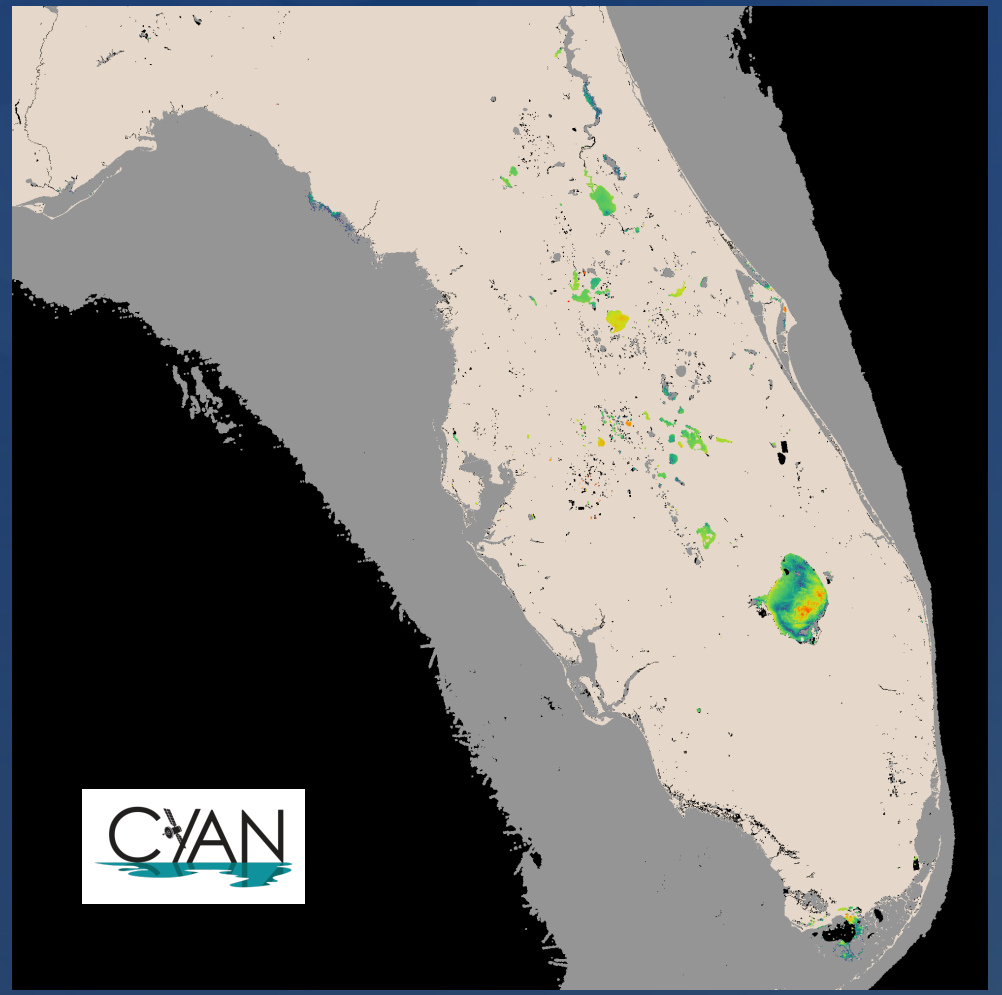
OLCI: Ocean and Land Colour Instrument

Response variable: Cyanobacterial abundance over a 7-day period

- Satellite-derived products increased spatial coverage of data
- Weekly composite CyAN imagery was used to estimate cyanobacterial abundance.



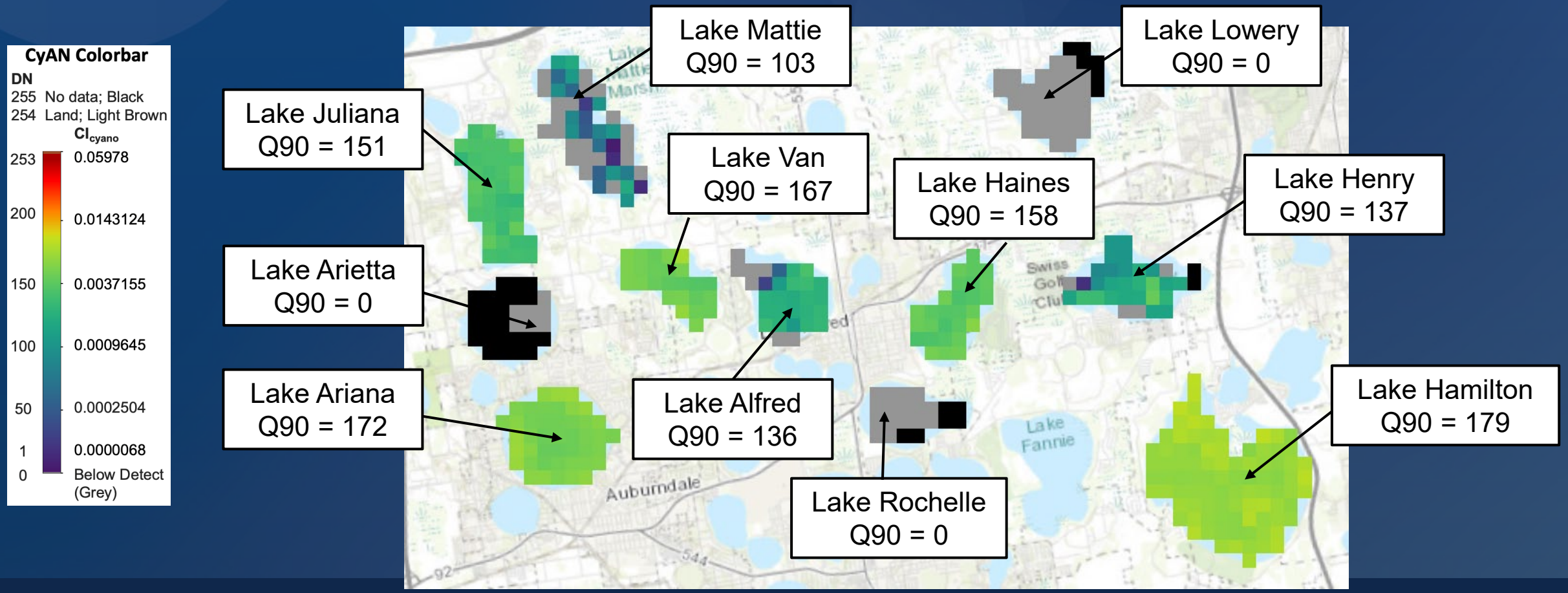
Daily CyAN Images



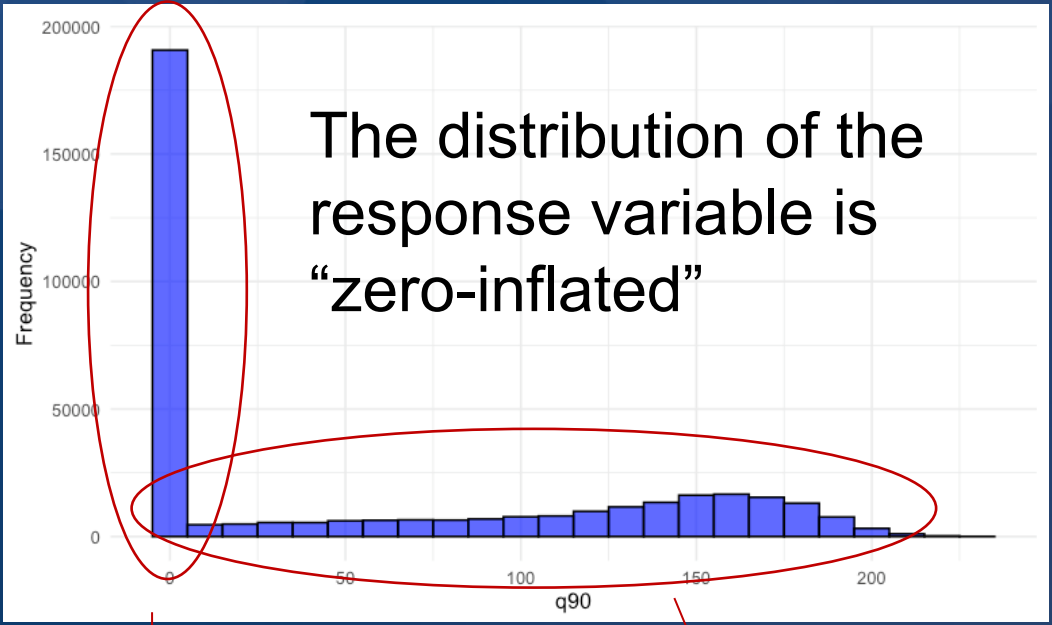
Weekly Maximum Composite

For each lake/zone, we computed the 90th percentile (Q90) of the weekly maximum composite CyAN value.

- **Definition of Q90**: 90% of the valid pixels in the lake/zone have a value equal to or lower than the Q90 value, i.e., Q90 is “almost the maximum”



Modeling Approach: Building a regression model to predict the CyAN Q90 value



The **zero-adjusted negative binomial model** has 2 components:

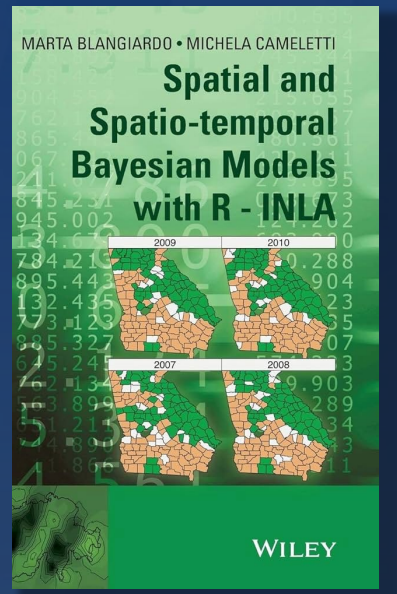
1. Binomial component (Is Q90 zero?)
2. Negative binomial component (if non-zero, predict the Q90 value)

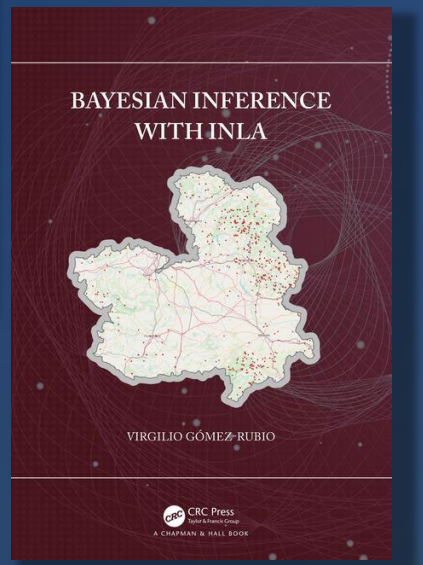
This part of the data is modeled using a logistic regression model

This part of the data is modeled using a negative binomial regression model

We used the Integrated Nested Laplace Approximation (**INLA**) Bayesian modeling R package to build our predictive model.

- INLA is a method for Bayesian inference.
- Key benefits of INLA:
  - Faster than other sampling methods (like MCMC)
  - The R-INLA package has lots of handy functionality for specifying and fitting models
  - The Bayesian approach allows us to easily compute confidence intervals for the predictions





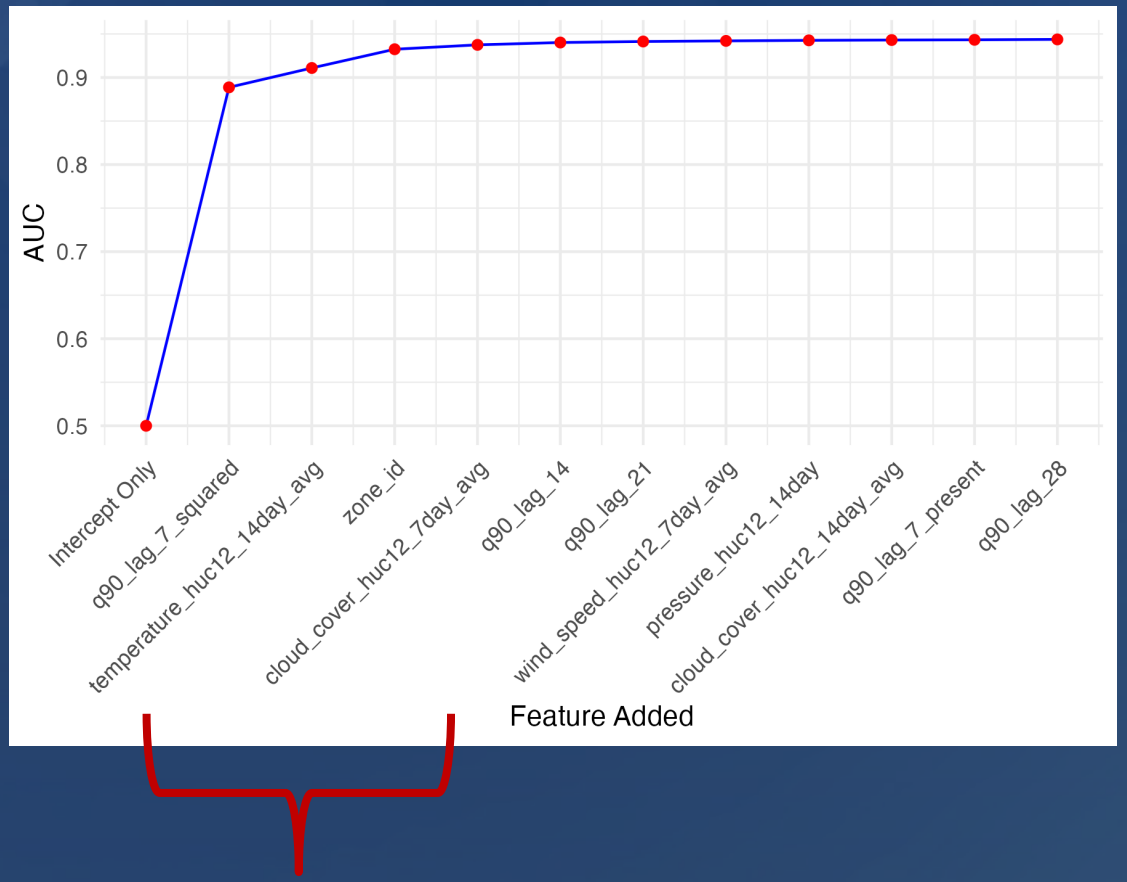
www.r-inla.org

Predictor variables considered

Category	Variable	Source
Physical variables	Area	NHD
	Mean depth	HydroLakes; Bathymetry data
	Lake classification	FDEP IWR
	Level 4 ecozone	EPA Ecozones
Meteorological variables	Daily wind speed	RTMA
	7-day average wind speed	RTMA
	14-day average wind speed	RTMA
	Daily wind direction	RTMA
	Daily pressure	RTMA
	Daily cloud cover	RTMA
	7-day average cloud cover	RTMA
	14-day average cloud cover	RTMA
	14-day average temperature	RTMA
	14-day total precipitation	MRMS
CyAN variables	7-day lagged Q90	CyAN
	14-day lagged Q90	CyAN
	21-day lagged Q90	CyAN
	7-day lagged Q90^2	CyAN
Seasonal variables	Cyclic date	
Fixed effects	Lake/zone-specific intercept	

Selecting predictor variables using a stepwise process

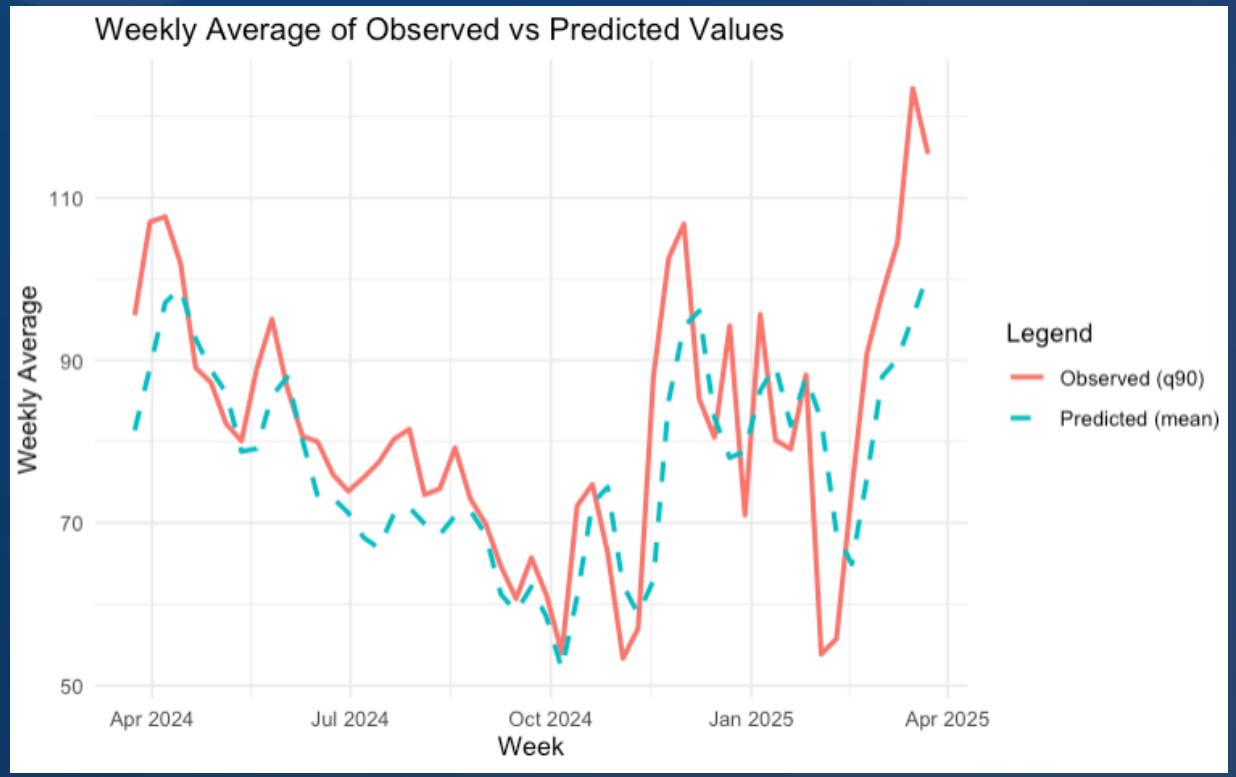
- Forward stepwise feature selection: Add variables one at a time, sorted by their impact on performance.
  - Measure performance using **AUC (**for binary) or **AIC** (for continuous)
  - The key benefit of forward stepwise feature selection is that we can identify the *point of diminishing returns*



Here, we would select the first ~4 features for inclusion in the model

Summary of the Selected Model

Predictors (binary component)	Predictor (continuous component)	Mean Absolute Error	Bias
Intercept, Zone ID, 7-day CyAN Q90 lag squared, 14-day average temperature, 7-day average cloud cover	Intercept, Lake ID, 14-day CyAN Q90 lags, 7-day Q90 lag squared, cyclic date	21.7 (14.3 when Q90>150)	-3.7



Model evaluation setup:

- Train period: April 2021-April 2024
- Test period: April 2024-April 2025

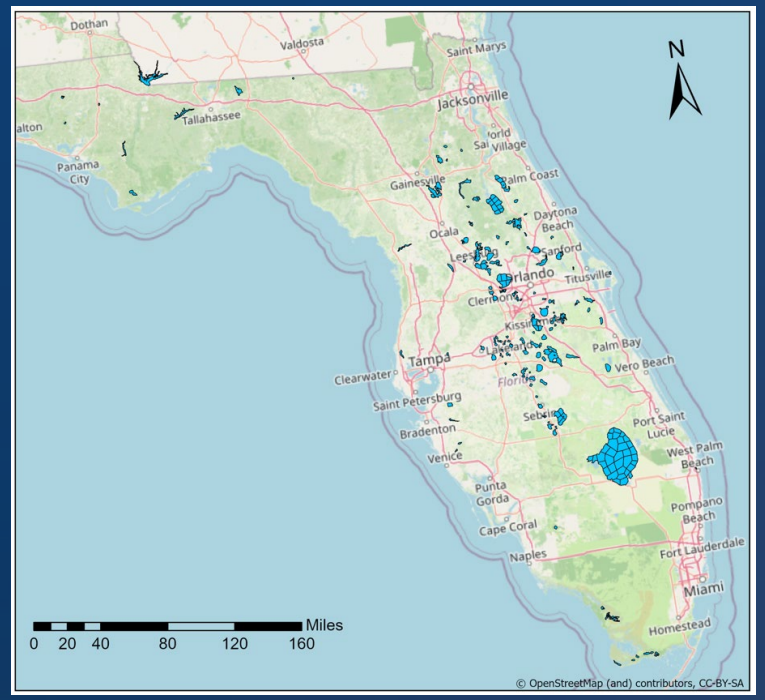
Model Accuracy for bloom/no-bloom

- We can establish a cutoff value (e.g., DN > 170) and assess the model's accuracy for predicting an exceeding value
- Using a cutoff of chlorophyll-a > 40 µg/L, **the accuracy of the model is 95%**

Developing a Separate Model for Lake Okeechobee

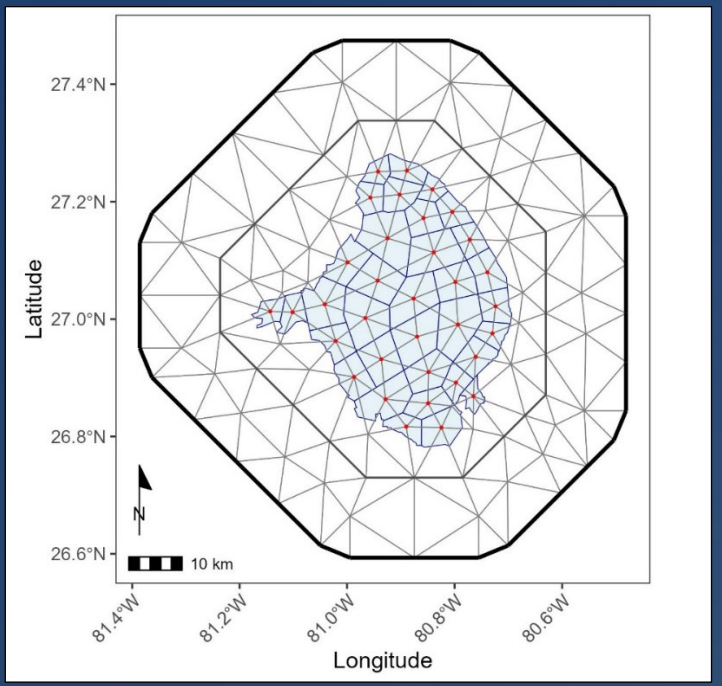
- Lake Okeechobee had many more zones than the other large lakes
- We included an additional predictor variable (spatial field or “mesh”) to account for spatial correlations between zones

“Florida Lakes Model”

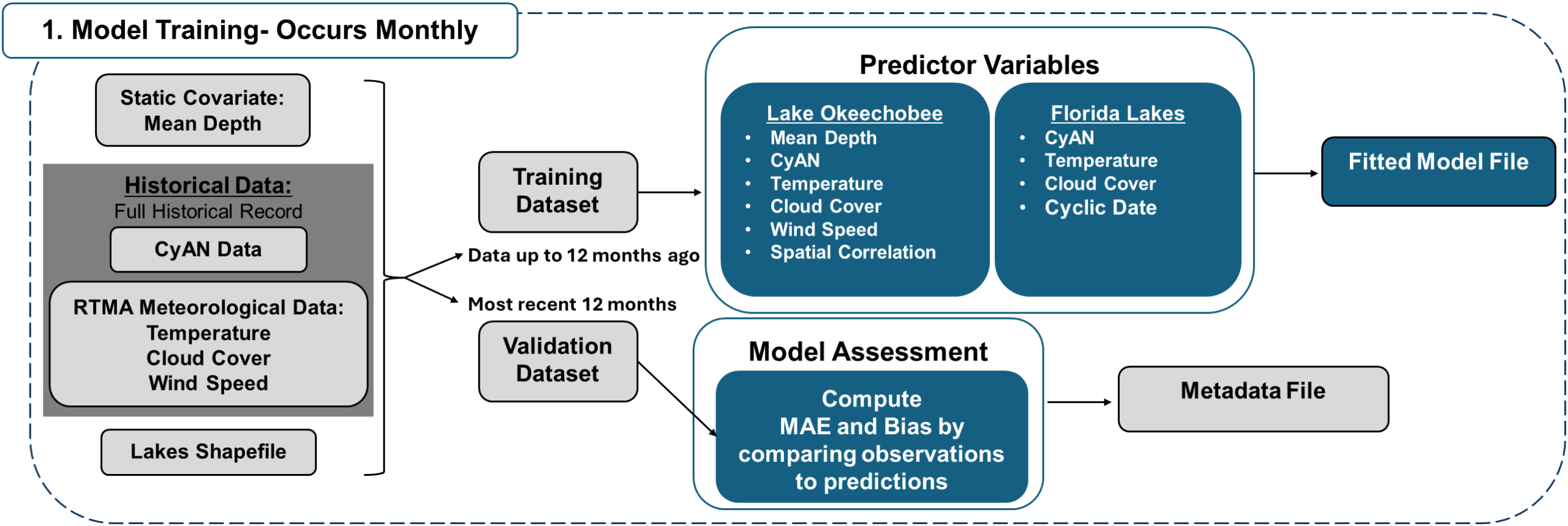


Lake	Number of Zones
Lake Okeechobee	34
Lake George	9
Lake Apopka	5
Lake Istokpoga	5
Crescent Lake	4
Lake Kissimmee	4
Lake Harris	3
Lake Tohopekaliga	3

“Lake Okeechobee Model”

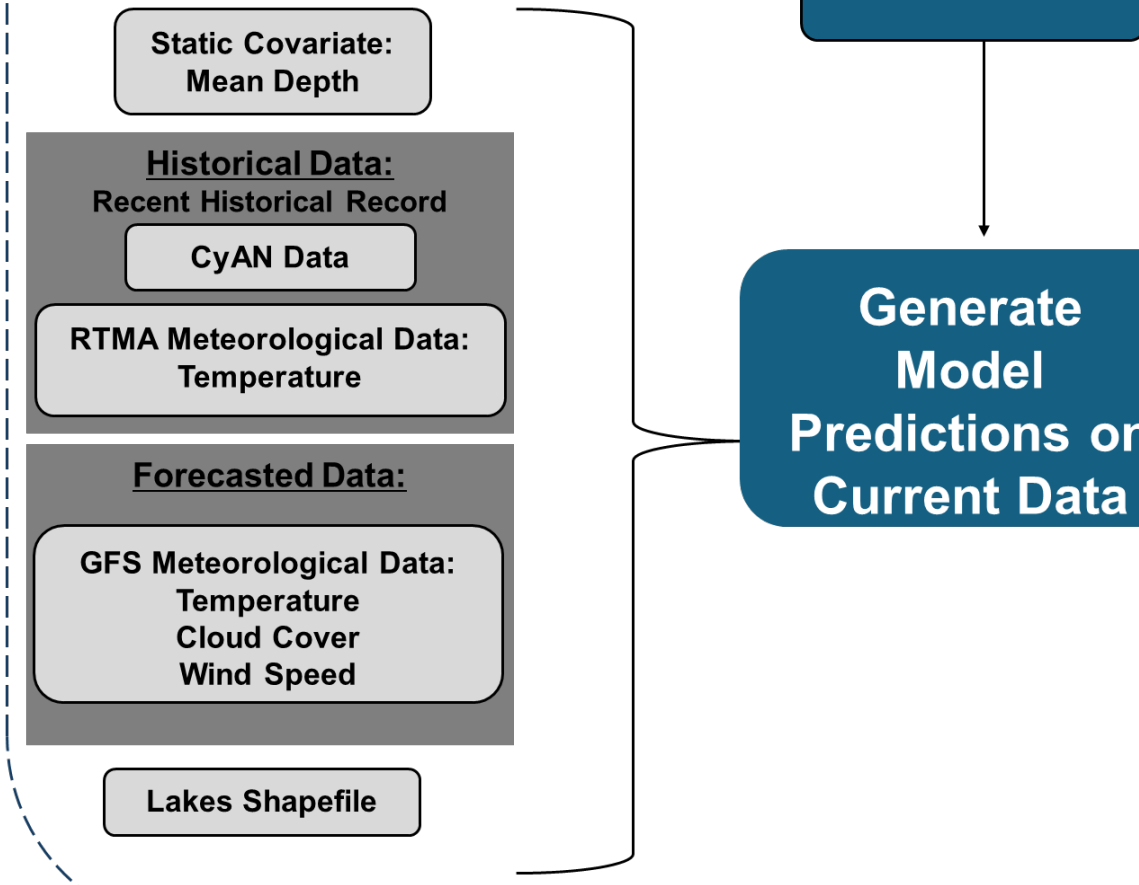


Operational Workflow: Model Training



Operational Workflow: Model Predictions

2. Issue Predictions- Occurs Daily



Fitted Model File

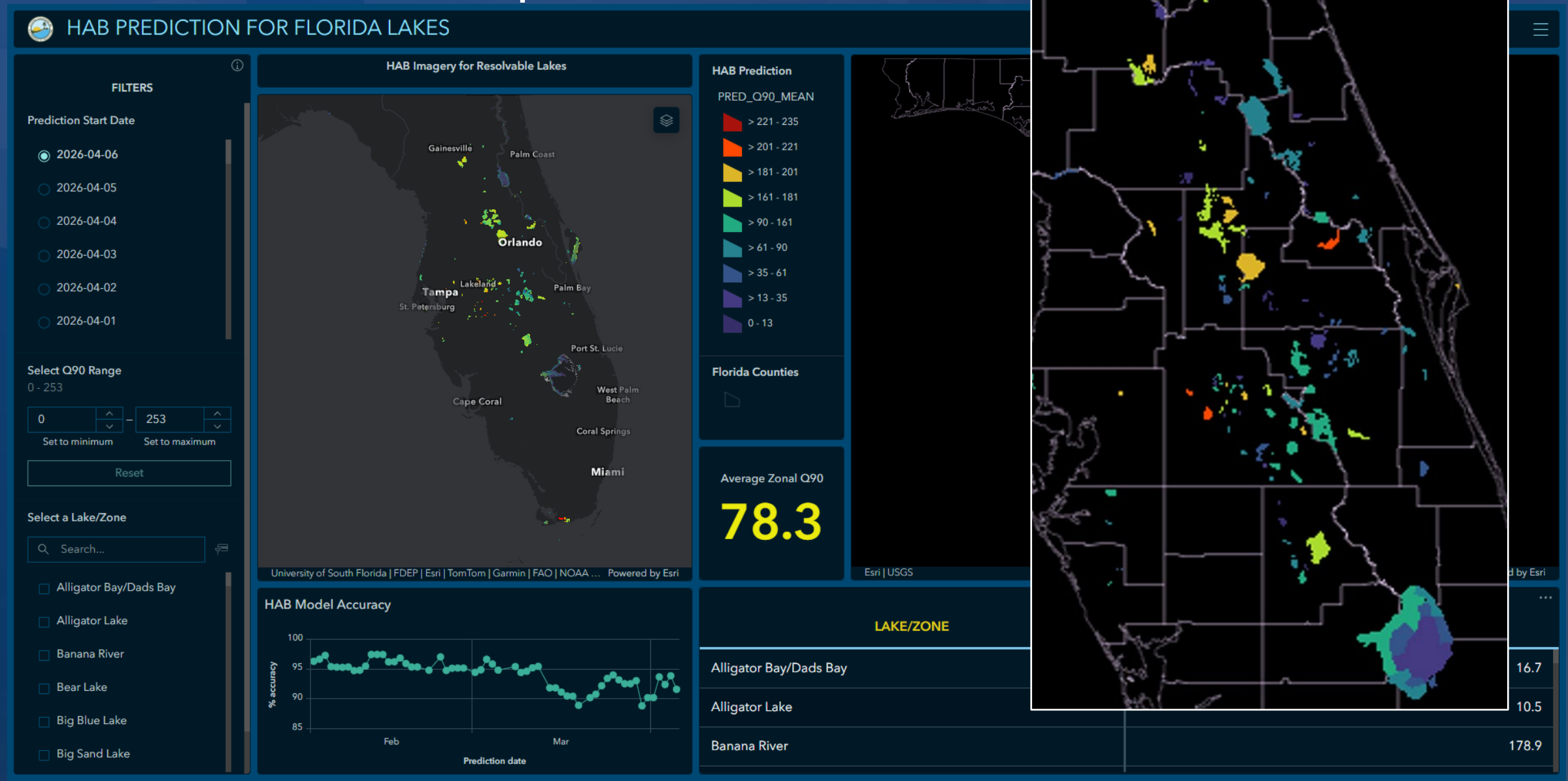
Generate Model Predictions on Current Data

Results File

ID	Name	Prediction Start Date	Prediction End Date	Predicted Q90	Issue Date
3309000	Georges Lake	6/12/2025	6/18/2025	43.20	2025-06-12T15:31:34Z
3312000	Lake Lowery	6/12/2025	6/18/2025	1.38	2025-06-12T15:31:34Z
3313000	Lake Griffin	6/12/2025	6/18/2025	197.15	2025-06-12T15:31:34Z
3314000	Lake Minneola	6/12/2025	6/18/2025	2.99	2025-06-12T15:31:34Z
3315000	Lake Ashby	6/12/2025	6/18/2025	6.25	2025-06-12T15:31:34Z
3316001	Lake Harris Zone 1	6/12/2025	6/18/2025	153.44	2025-06-12T15:31:34Z
3316002	Lake Harris Zone 2	6/12/2025	6/18/2025	138.94	2025-06-12T15:31:34Z
3316003	Lake Harris Zone 3	6/12/2025	6/18/2025	144.74	2025-06-12T15:31:34Z

Run Log File

Florida DEP Model Graphical User Interface



Model Strengths

- **Predicts a Continuous Variable**
- **Two-stage model separately captures bloom occurrence and intensity**
- **Automated workflow for near-real time forecasting**

Model Limitations

- **Reduced sensitivity to short-term change in bloom dynamics**
- **Dependance on lagged predictor variables**
- **Validation with CyAN imagery rather than field data**

Thank you

- Contact Us
  - Jonathan Holt (jonathanholt@rti.org)
  - Natalie Reynolds (nreynolds@rti.org)

This work was conducted by RTI as under subaward to SAS Institute and funded by the Florida Department of Environmental Protection.