On the Sensitivity of a Machine Learning-Based Model to Predict Chlorophyll-α Using Multi-Media Modeling Environmental Predictors

C. Feng Chang[1], M. Astitha[1], V. Garcia[2], C. Tang[2], P. Vlahos[3], D. Wanik[4], J. Bash[2]

[1]Deparment of Civil and Environmental Engineering, University of Connecticut
[2]National Exposure Research Laboratory, Office of Research and Development, US EPA
[3]Department of Marine Sciences, University of Connecticut
[4]Department of Operations and Information Management, University of Connecticut

19th ANNUAL
CMAS Conference
Oct. 26-30 | Virtual

Email: christina.feng_chang@uconn.edu
Group website: airmg.uconn.edu

University of Connecticut
School of Engineering

Scope and Objectives

SCOPE: Use **multi-media modeling** and **machine learning (ML)** to assess algal blooms

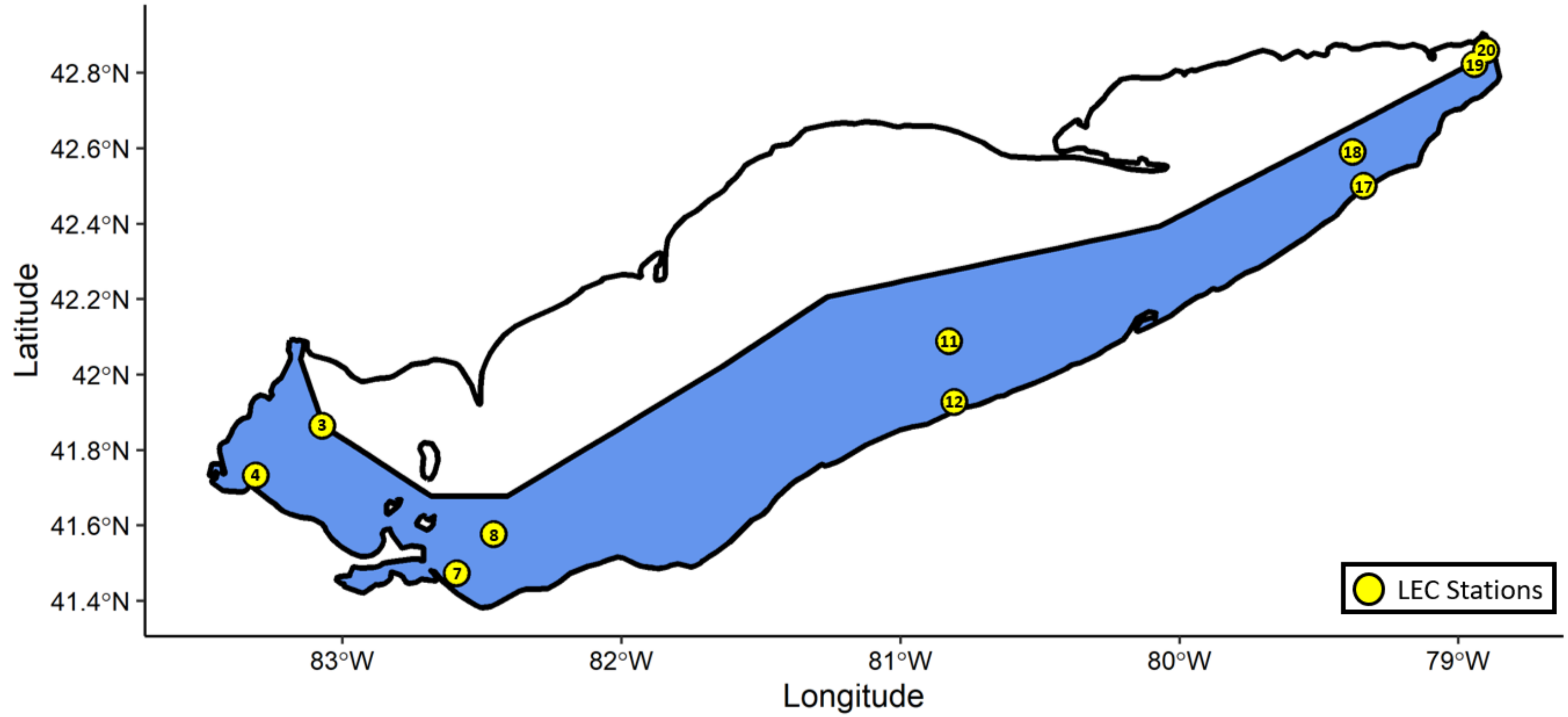
1. Developed a ML-based methodology that integrated modeled and observed data to:
   - Identify and evaluate important environmental parameters
   - Predict chlor-α
2. **Examine the sensitivity of the model by varying input modeled data (meteorological, air quality, hydrological, and agricultural) through sensitivity tests**

MERIS Sensor, European Space Agency (ESA) Envisat, 2011

University of Connecticut School of Engineering

MODEL DATA: Observed Variable

United States part of Lake Erie (2002-2012)



- In-situ chlor-α measurements provided by:
    - Lake Erie Committee Forage Task Group (LEC FTG)
- Chlor-α measurements are seasonally averaged (May to September)

University of Connecticut
School of Engineering

MODEL DATA: Modeled Variables

Numerical Prediction Models

12-km grid spacing

Explanatory Variables

Meteorology
(WRF)
air temp, wind, precipitation, radiation

Hydrology
(VIC)
soil moisture, water flow, evaporation, water temp

Agriculture Management
(USDA EPIC)
fertilizer applications

Air Quality
(EPA CMAQ-Bidi)
atmospheric N deposition

Lake Water Quality Assessment and Prediction
Machine Learning Model
(Random Forest)

Target Variable
Chlor-α

Water Quality Indicator Observations
(Chlorophyll-α)

MODEL DATA: Modeled Variables

3 types of variables

- ***Static*** – indicate location of measurement in lake
- ***Point*** – paired to the closest model grid point to each chlor-α station
- ***Watershed (WS)*** – aggregated daily values for all grids in the HUC-8 watershed draining into the lake

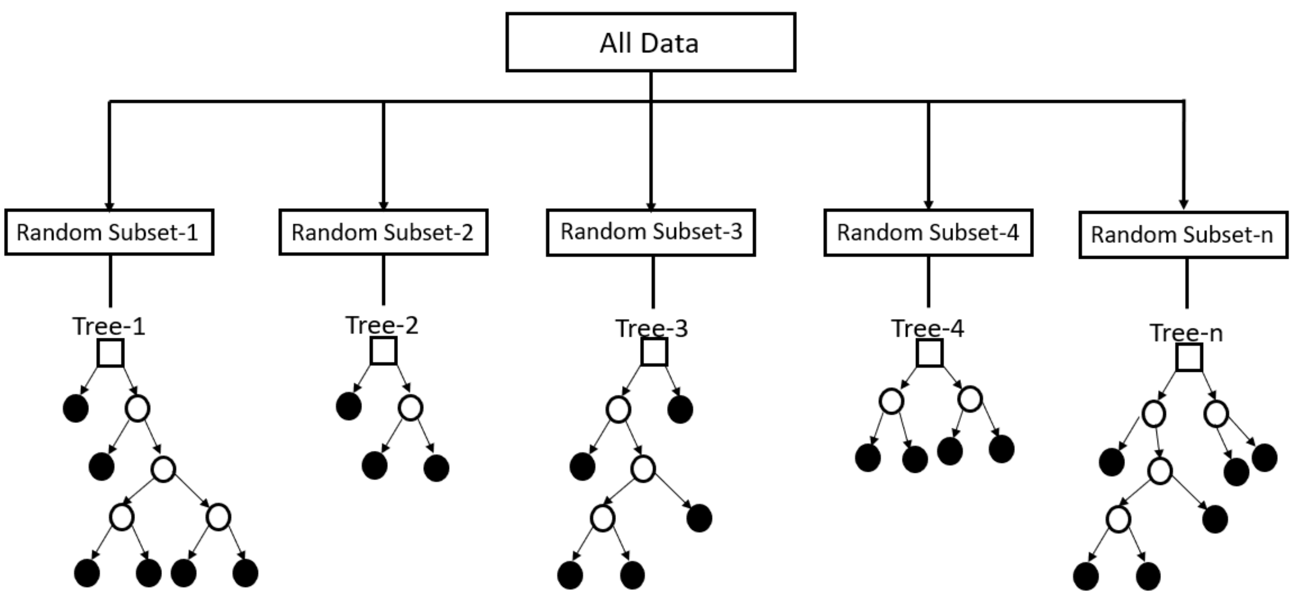
29 variables evaluated

Explanatory Variables	Units	Model
Latitude (static variable)	degrees (°)	
Longitude (static variable)	degrees (°)	
Max_Radiation (Point)	W/m^2	WRF
Tmax (Point)	℃	WRF
Tmax_Days_Above_25 (Point)	days	WRF
Precipitation (WS)	mm	WRF
Avg_Wind (Point)	m/s	WRF
Dry_Oxidized_ND (WS)	g/ha	CMAQ
Dry_Reduced_ND (WS)	g/ha	CMAQ
Wet_Oxidized_ND (WS)	g/ha	CMAQ
Wet_Reduced_ND (WS)	g/ha	CMAQ
Wet_Organic_ND (WS)	g/ha	CMAQ
Evaporation (Point)	kg/m^2	VIC
Water Flow (WS)	cfs	VIC
Soil moisture Layer 1 (0-10 cm) (Point)	kg/m^2	VIC
Soil moisture Layer 2 (10-40 cm) (Point)	kg/m^2	VIC
Soil moisture Layer 3 (40-150 cm) (Point)	kg/m^2	VIC
Water_Temp_C (Point)	℃	VIC
Water_Temp_Days_Above_25 (Point)	days	VIC
Layer1 N-NO3 (Nitrate) Application Rate (WS)	tons	EPIC
Layer1 N-NH3 (Ammonia) Application Rate (WS)	tons	EPIC
Layer1 ON (Organic N) Application Rate (WS)	tons	EPIC
Layer1 MP (Mineralized P) Application Rate (WS)	tons	EPIC
Layer1 OP (Organic P) Application Rate (WS)	tons	EPIC
Layer2 N-NO3 (Nitrate) Application Rate (WS)	tons	EPIC
Layer2 N-NH3 (Ammonia) Application Rate (WS)	tons	EPIC
Layer2 ON (Organic N) Application Rate (WS)	tons	EPIC
Layer2 MP (Mineralized P) Application Rate (WS)	tons	EPIC
Layer2 OP (Organic P) Application Rate (WS)	tons	EPIC

METHODOLOGY

- **Step 1:** Train and validate the random forest (RF) model with all explanatory variables.
- **Step 2:** Examine performance of RF model through 10-fold CV and evaluate importance of top explanatory variables through accumulated local effect (ALE) plots.
- **Step 3:** Evaluate the influence of each set of parameters through sensitivity tests by varying EPIC, VIC, WRF, and CMAQ inputs into the RF model.

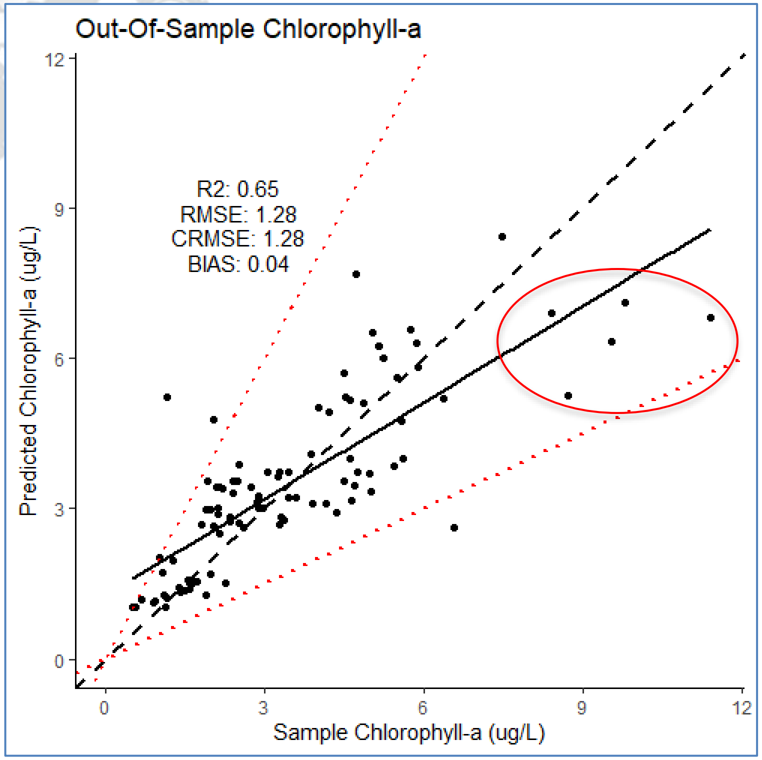
Simplified RF Diagram



University of Connecticut
School of Engineering

Results: Prediction of chlor-α

Prediction of Chlor-a



Eutrophic Threshold:
Chlor-α > 5μg/L

Contingency Table

Chlor-α >5 μg/L		OBSERVATIONS	
		YES	NO
MODEL	YES	14	7
	NO	5	70

Total points = 96

PC = 87.5%

POD_1 (Chlor-α>5 μg/L) = 73.7%

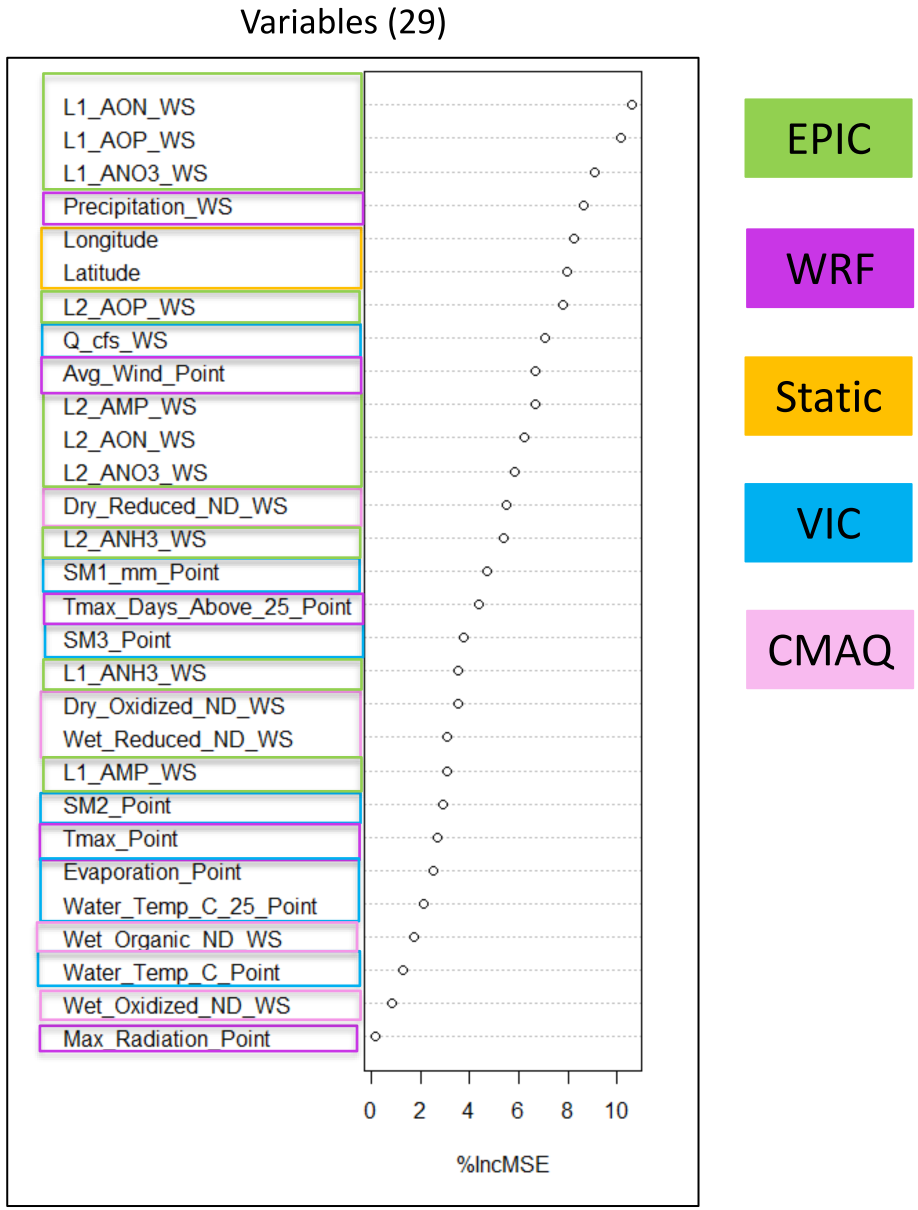
POD_2 (Chlor-α≤5 μg/L) = 90.9%

- 65% of variance in chlor-α measurements is explained by the RF model
- 94.8% of the model's predictions are within a factor of 2 of the obs

- Eutrophic conditions are identified 73.7% of the time
- Detection of eutrophic vs. non-eutrophic conditions is 87.5%

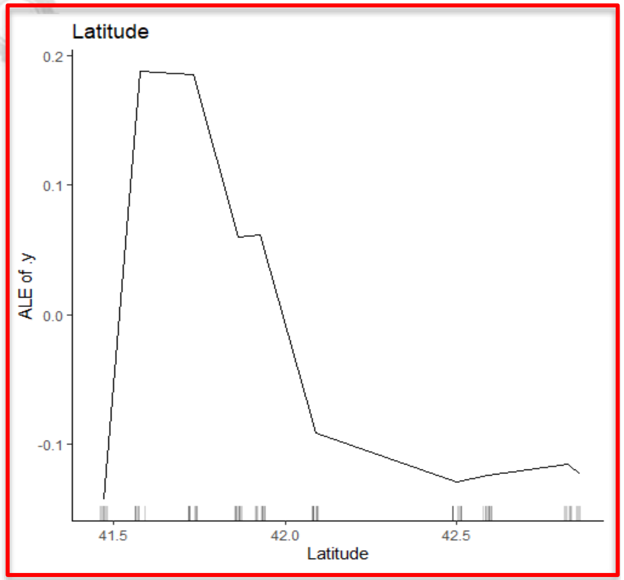
University of Connecticut
School of Engineering

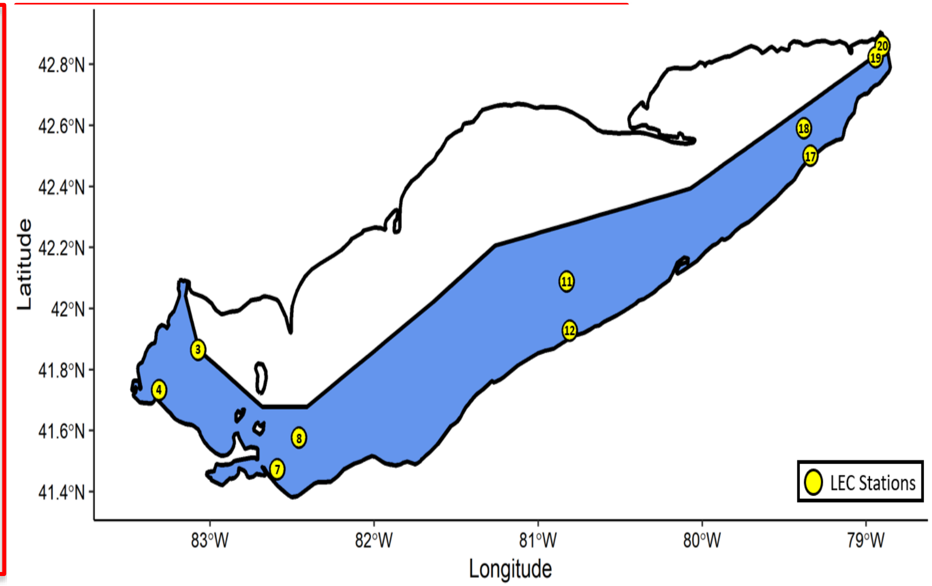
Results: Variable Importance



Discussion: Variable Effects

Accumulated Local Effect (ALE) Plots





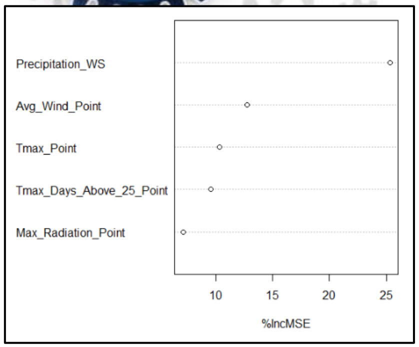
- Higher chlor-α concentrations from west to east of the lake **(FTG LEC, 2019)**

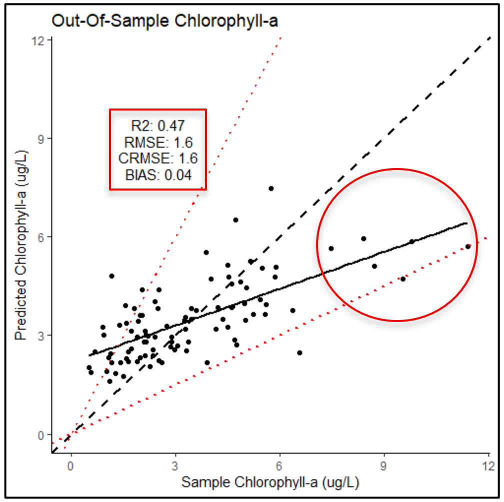
- An increase in precipitation leads to an increase in chlor-α
  - Precipitation and river discharge increase nutrient loads delivered to the lake **(Stow et al., 2015; EPA GLNPO U.S. Action Plan For Lake Erie, 2018)**

University of Connecticut
School of Engineering

Discussion: Sensitivity Tests

WRF Only

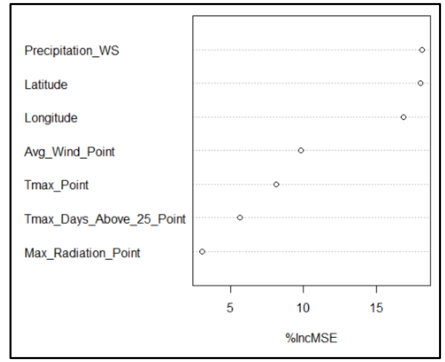


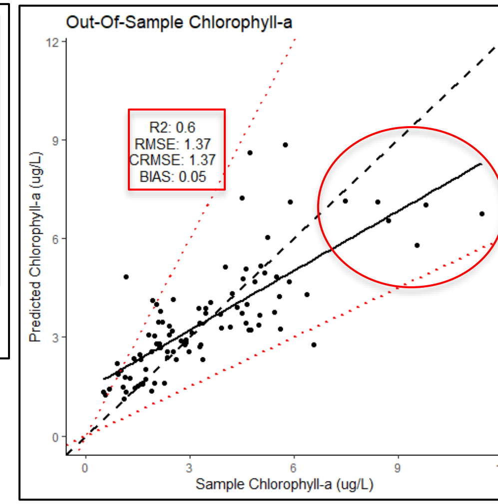


Chlor-α >5 μg/L		OBSERVATIONS	
		YES	NO
MODEL	YES	9	3
	NO	10	74

Total points = 96
PC = 86.5%
POD_1 (Chlor-α>5 μg/L) = 47.4%
POD_2 (Chlor-α≤5 μg/L) = 96.1%

WRF and Static





Chlor-α >5 μg/L		OBSERVATIONS	
		YES	NO
MODEL	YES	10	4
	NO	9	73

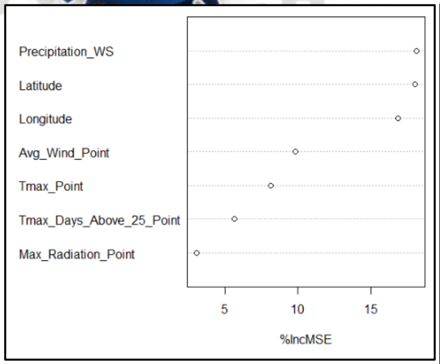
Total points = 96
PC = 86.5%
POD_1 (Chlor-α>5 μg/L) = 52.6%
POD_2 (Chlor-α≤5 μg/L) = 94.8%

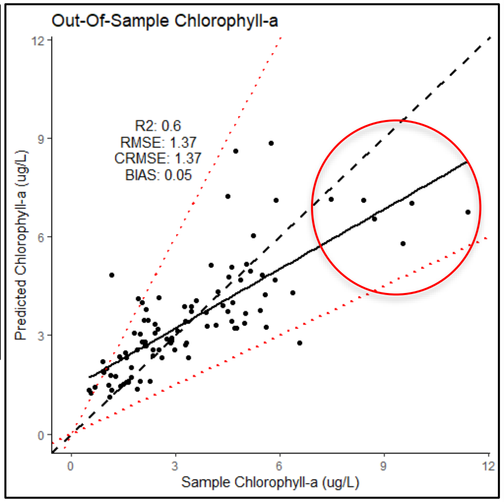
- Spatial variables latitude and longitude are significantly important

University of Connecticut
School of Engineering

Discussion: Sensitivity Tests

WRF and Static





Chlor-α >5 µg/L		OBSERVATIONS	
		YES	NO
MODEL	YES	10	4
	NO	9	73

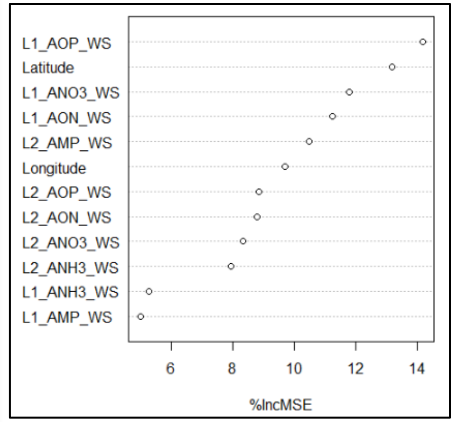
Total points = 96

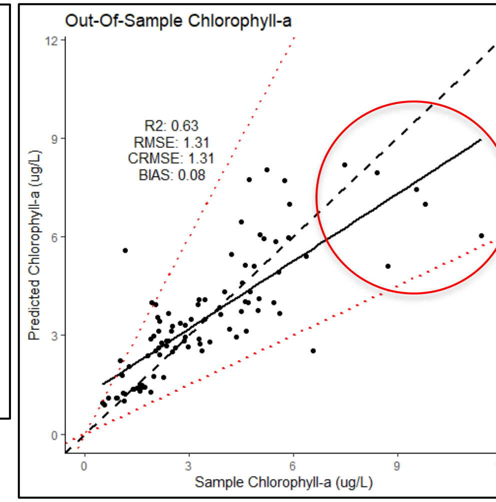
PC = 86.5%

POD_1 (Chlor-α>5 µg/L) = 52.6%

POD_2 (Chlor-α≤5 µg/L) = 94.8%

EPIC and Static





Chlor-α >5 µg/L		OBSERVATIONS	
		YES	NO
MODEL	YES	14	6
	NO	5	71

Total points = 96

PC = 88.5%

POD_1 (Chlor-α>5 µg/L) = 73.7%

POD_2 (Chlor-α≤5 µg/L) = 92.2%

- Meteorological variables alone perform poorly for predicting high chlor-α values
- Fertilizer application variables are influential for predicting high chlor-α values
  - Accelerated eutrophication from human activities links to harmful algal blooms **(Watson et al., 2016; Anderson et al. 2002)**
  - Joint efforts from the U.S. and Canada to reduce P loadings **(Canada-Ontario Lake Erie Action Plan, 2018)**

Limitations

- Lack of high chlor-α observations for the RF model to train on
- No lake hydrodynamic information (e.g., lake thermal structure, water motions)
- Wastewater discharges from industrial and municipal sources were not included
- No information on the Canadian portion of Lake Erie (US contributes to 84% of total P loads to Lake Erie) ***(Canada-Ontario Lake Erie, 2018)***

University of Connecticut
School of Engineering

SUMMARY and FUTURE WORK

- The model has been improved with updated versions of VIC, WRF, and EPIC
- Results are promising but more quantitative assessment is necessary
- The model ranked 29 influential variables conducive to a successful prediction of chlor-α: (1) N and P fertilizer applications, (2) location, (3) meteorology, hydrology, and air quality
- Expand study period from 11-years (2002-2012) to 16-years (2002-2017) with new CMAQ simulations
- Try different machine learning approaches
- Given sufficient record of data, the predictive tool can be applied to other lakes and coastal locations to study other water quality indicators
- Build a predictive tool capable of providing water quality forecasts

References

- Anderson, D.M., Glibert, P.M. & Burkholder, J.M. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries* 25, 704–726. **2002**.
- Canada-Ontario Lake Erie Action Plan – Partnering on Achieving Phosphorus Loading Reductions to Lake Erie from Canadian Sources. Environment and Climate Change Canada and the Ontario Ministry of the Environment and Climate Change, Ontario, Canada, **2018**.
- Forage Task Group. Report of the Lake Erie Forage Task Group, March **2019**. Presented to the Standing Technical Committee, Lake Erie Committee of the Great Lakes Fishery Commission, Ann Arbor, Michigan, USA.
- Stow, C.A.; Y. Cha; L.T. Johnson; R. Confesor; R.P. Richards. **2015.** Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie. Environmental Science & Technology, 49: 3392-3400.
- U.S. Action Plan for Lake Erie. U.S. Environmental Protection Agency, Great Lakes National Program office. **2018**.
- Watson, S.B.; Miller, C.; Arhonditsis, G.; Boyer, G.L.; Carmichael, W.; Charlton, M.N.; Confesor, R.; Depew, D.C.; Hook, T.O.; Ludsin, S.A.; Matisoff, G.; McElmurry, S.P.; Murray, M.W.; Richards, R.P; Rao, Y.R.; Steffen, M.M.; Wilhelm, S.W. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae* 56, 44-66. **2016**.

Acknowledgements

- Special thanks to Jesse Bash for providing the new EPIC and WRF simulations; Chunling Tang for running new VIC simulations; and Valerie Garcia and Marina Astitha for the constant guidance and support.

- We would also like to thank James Markham and Patrick Kocovsky from the Lake Erie Forage Task Group for providing valuable information and guidance on utilizing the LEC data.

Questions:
Email: christina.feng_chang@uconn.edu

University of Connecticut
School of Engineering