Using Machine Learning to Assess Parameters Associated with Harmful Algal Blooms for Lake Erie

C. Feng Chang[1], M. Astitha[1], V. Garcia[2], C. Tang[2], P. Vlahos[3], D. Wanik[4], J. Yan[5]

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

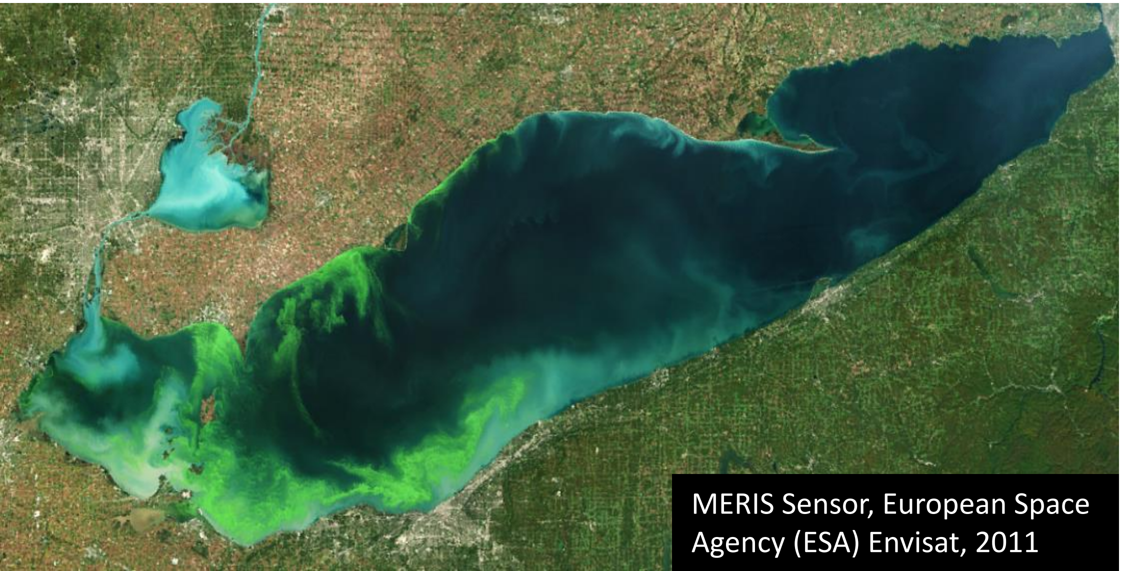
Email: christina.feng_chang@uconn.edu
Group website: airmg.uconn.edu

University of Connecticut
School of Engineering

Scope and Objectives

SCOPE: ***Investigate*** **and** ***predict*** **seasonal growth of algal blooms using chlorophyll-α (chlor-α) as a proxy**

- Predict chlor-α concentrations
- Identify and evaluate the importance of environmental parameters in these predictions
- Analyze and assess the contribution of atmospheric nitrogen deposition on chlor-α



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:
  - Great Lakes National Program Office's Great Lakes Environmental Database System (GLNPO GLENDA)
  - Lake Erie Committee Forage Task Group (LEC FTG)
- Chlor-α measurements were seasonally averaged (April to September)

University of Connecticut
School of Engineering

MODEL DATA: Modeled Variables

Numerical Prediction Models

12-km grid spacing

Explanatory Variables

Meteorology
(WRF)

- Additional 5 variables for each modeled variable representing a time lag of 1 to 5 days prior to sample collection date
- Three types of variables:
  - ***Point*** – paired to closest model grid point to each chlor-α station
  - ***Watershed (WS)*** – average daily values for all grids in the HUC-8 watershed draining into lake
  - ***Static*** – indicate location

Over 250 variables evaluated

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

Explanatory Variables	Units	Model
Latitude (static variable)	degrees (°)	
Longitude (static variable)	degrees (°)	
Radiation (Point)	W/m^2	WRF
Taverage (Point, WS)	°C	WRF
Precipitation (Point, WS)	mm	WRF
R_humidity (Point)		WRF
Windspeed (Point)	m/s	WRF
Dry_Oxidized_N (Point, WS)	kg/ha	CMAQ
Dry_Reduced_N (Point, WS)	kg/ha	CMAQ
Wet_Oxidized_N (Point, WS)	kg/ha	CMAQ
Wet_Reduced_N (Point, WS)	kg/ha	CMAQ
Wet_Organic_N (Point, WS)	kg/ha	CMAQ
Evapotranspiration (Point)	mm	VIC
Water Flow (WS)	Cfs	VIC
Soil moisture Layer 1 (0-10 cm) (Point)	mm	VIC
Soil moisture Layer 2 (10-40 cm) (Point)	mm	VIC
Soil moisture Layer 3 (40-150 cm) (Point)	mm	VIC
Water_Temp_C (Point)	°C	VIC
surface runoff (WS)	Mm	EPIC
soil loss from water erosion (WS)	ton/ha	EPIC
N loss with sediment (WS)	kg/ha	EPIC
P loss with sediment (WS)	kg/ha	EPIC
nitrate loss in surface runoff (WS)	kg/ha	EPIC
labile P loss in surface runoff (WS)	kg/ha	EPIC
N in subsurface flow (WS)	kg/ha	EPIC
soluble N in drainage outflow (WS)	kg/ha	EPIC
soluble P loss through drainage system (WS)	kg/ha	EPIC
Layer1 N-NO3 (Nitrate) Application Rate (WS)	kg/ha	EPIC
Layer1 N-NH3 (Ammonia) Application Rate (WS)	kg/ha	EPIC
Layer1 ON (Organic N) Application Rate (WS)	kg/ha	EPIC
Layer1 MP (Mineralized P) Application Rate (WS)	kg/ha	EPIC
Layer1 OP (Organic P) Application Rate (WS)	kg/ha	EPIC
Layer2 N-NO3 (Nitrate) Application Rate (WS)	kg/ha	EPIC
Layer2 N-NH3 (Ammonia) Application Rate (WS)	kg/ha	EPIC
Layer2 ON (Organic N) Application Rate (WS)	kg/ha	EPIC

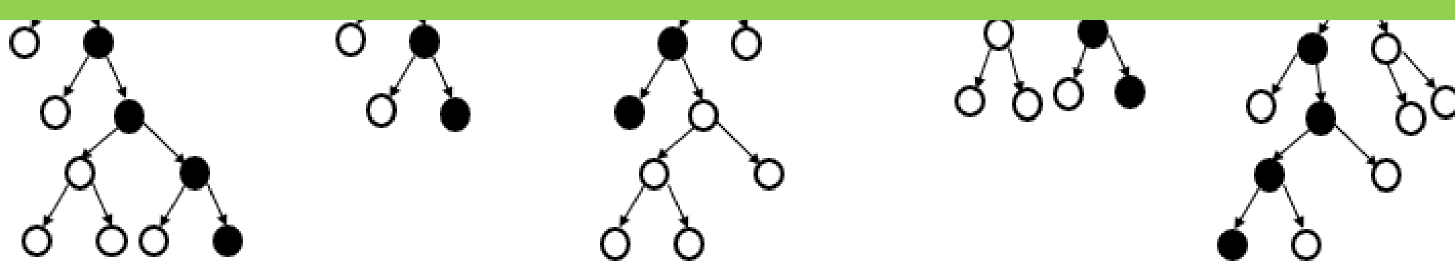
University of Connecticut
School of Engineering

METHODOLOGY

- Ra… s pre…
- Hy… no… pe…

Modeling Work Flow:

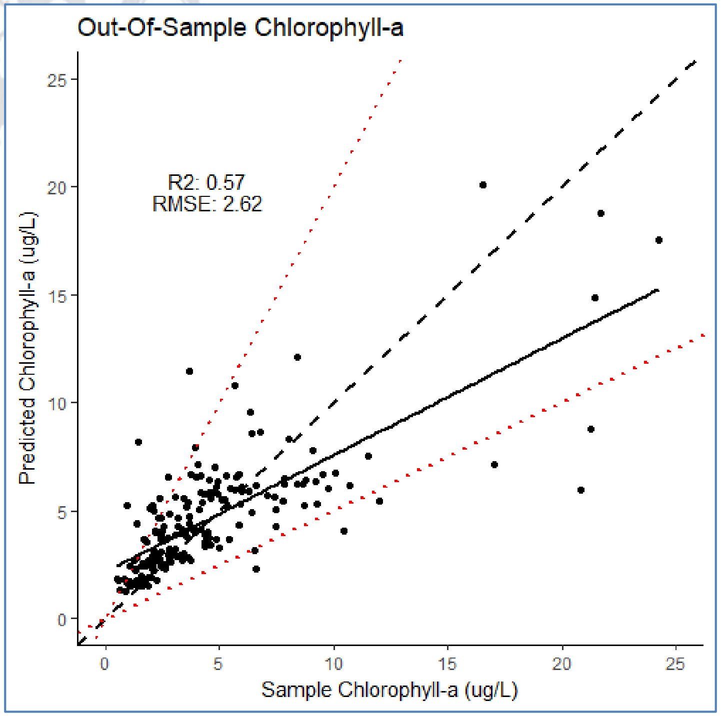
- **Step 1:** Train and validate RF model with all explanatory variables together with randomly generated variables (used to reduce noise).
  - 32 explanatory variables remain
- **Step 2:** Tune hyperparameters: mtry and ntree
- **Step 3:** Examine performance of the RF model through 10-fold CV and evaluate importance of top explanatory variables through accumulated local effect (ALE) plots
- **Step 4:** Test the approach using 2012 as an individual holdout year by creating a separate RF model using data from 2002-2011 to train and validate the model.



rsity of ecticut ineering

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	46	29
	NO	8	104

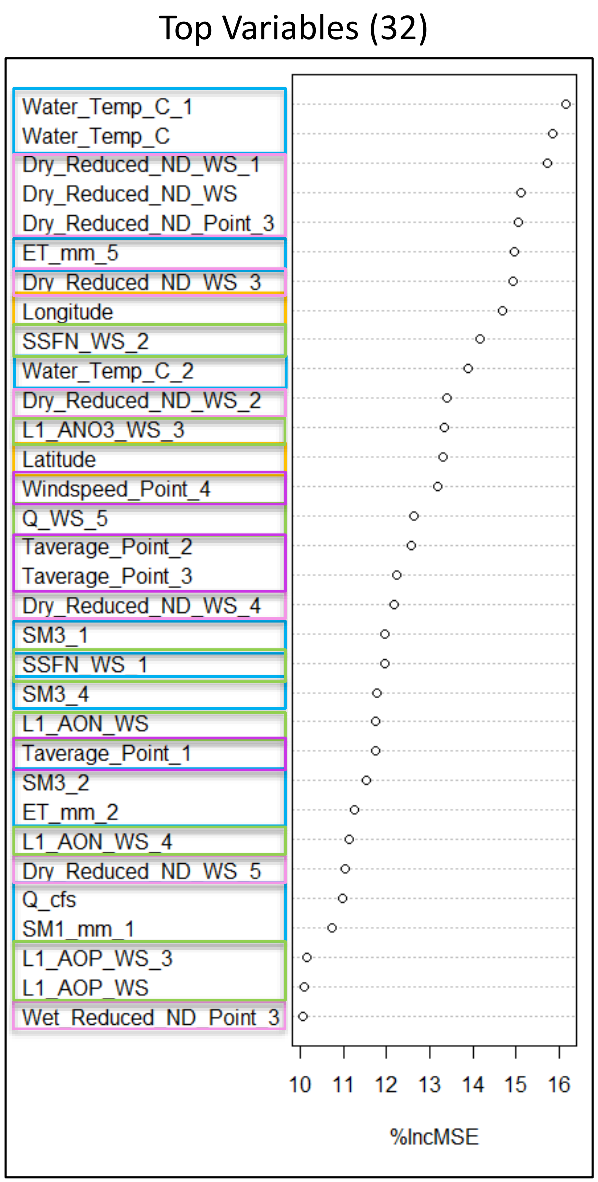
Total points = 187
PC = 80.2%
POD_1 (Chlor-α >5 μg/L) = 85.1%
POD_2 (Chlor-α ≤5 μg/L) = 78.2%

- Almost 60% of variance in chlor-α measurements is explained by the RF model
- 86.6% of the model's predictions are within a factor of 2 of the obs

- Eutrophic conditions are identified 85.1% of the time
- Detection of eutrophic vs. non-eutrophic conditions is 80.2%

University of Connecticut
School of Engineering

Results: Variable Importance



VIC

CMAQ

Static

EPIC

WRF

University of Connecticut
School of Engineering

Discussion: Deposition of Atmospheric N

- Findings are in line with recent studies identifying:

  - Atmosphere and tributaries in the US are shifting from NO_3-dominated environment to a NH_4-dominated environment (decreases in NO_x emissions but emissions of NH_3 and unregulated air pollutants are continuous) ***(Compton et al. 2011; Li et al. 2016; Newell et. al. 2019; Paerl et al. 2018)***

  - N loads in the Maumee River are shifting from oxidized to reduced forms of N on a seasonal basis ***(Newell et. al. 2019)***

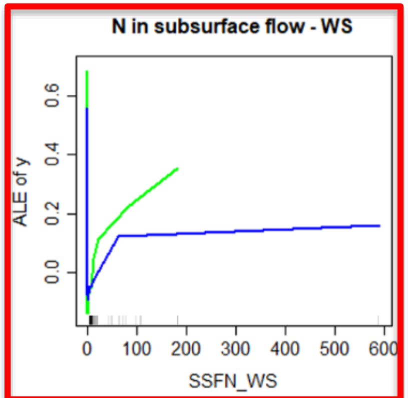
  - Strong association between reduced N loads and cyanobacterial growth ***(Newell et. al. 2019)***

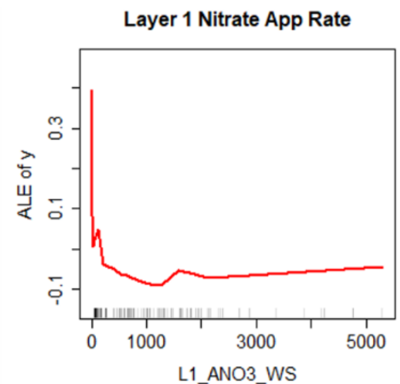
- CMAQ allows the inclusion of wet vs dry and oxidized vs. reduced atmospheric N deposition which have not been included in past HABs assessments.

- Dry c … 6 days
  -
- Dry a … a
  drast

University of Connecticut
School of Engineering

Discussion: Fertilizer Application





Layer 1 Organic N App Rate

Layer 1 Organic P App Rate

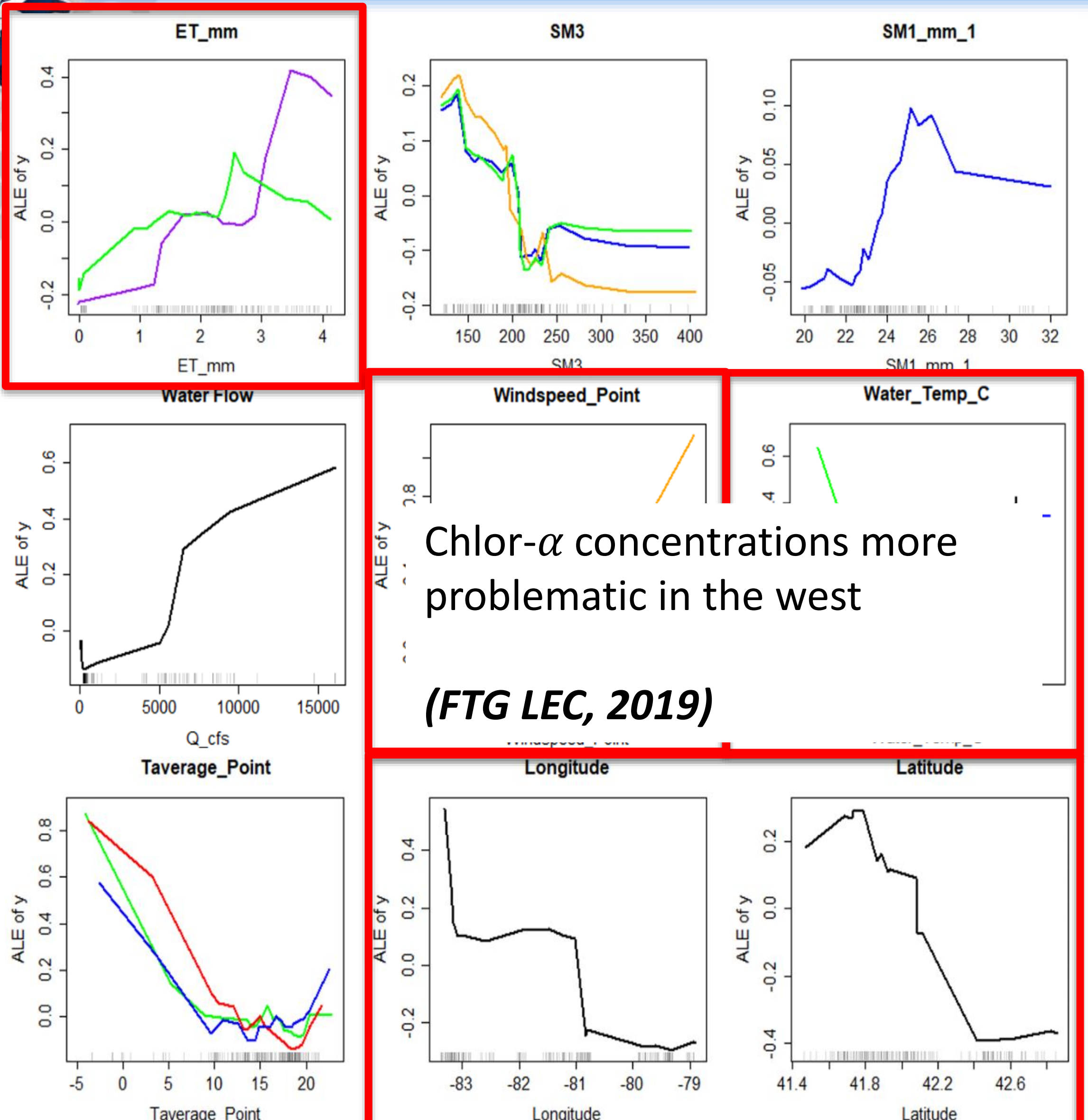
Surface Runoff

- N in subsurface flow increases, chlor-α increases
  - Ammonia from N fertilizers transforms to nitrate which easily leaches into groundwater and become a continuous source of nutrient into the lake and nearby streams
  - USGS indicates Lake Erie as an area of high risk for contamination of shallow groundwater by nitrate due to high N inputs (e.g., commercial fertilizer, atmospheric deposition, etc.) ***(U.S. Geological Survey Circular, 1999)***

- N a … nding on
- Su
  - Nutrients, sediments, and other pollutants entering the lake

University of Connecticut
School of Engineering

Discussion: Other Important Variables



Chlor-α concentrations more problematic in the west

(FTG LEC, 2019)

No lag – black, Lag 1 – blue, Lag 2 – green, Lag 3 – red, Lag 4 – orange, Lag 5 – purple

High ET increases chlor-α due to more stable and stagnant conditions

Spike in chlor-α when water temperatures >25°C, optimal temperature for cyanobacteria

Winds >5m/s drive resuspension events and carry nutrients stimulating initial algal growth

(Michalak et al. 2013)

University of Connecticut
School of Engineering

Limitations

- It is possible that it takes longer than 5 lag days for biological and chemical processes to occur
- 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)***
- No information on atmospheric deposition regarding P

University of Connecticut
School of Engineering

SUMMARY and FUTURE WORK

- The model identifies eutrophic conditions **over 85%** of the time
- Atmospheric deposition of reduced N plays an important role when it comes to chlor-α prediction
- The model identified 32 top influential variables conducive to a successful prediction of chlor-α: N and P fertilizer applications and both atmospheric and hydrologic conditions
- Given sufficient record of data, the predictive tool can be applied to other Great Lakes, other inland lakes, and coastal locations
- Similar approaches can be utilized to assess other water quality indicators: DO, total N, total P, and more

University of Connecticut
School of Engineering

References

- 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**.
- Compton, J. E.; Harrison, J. A.; Dennis, R. L.; Greaver, T. L.; Hill, B. H.; Jordan, S. J.; Walker, H.; Campbell, H. V. Ecosystem services altered by human changes in the nitrogen cycle: a new perspective for US decision making. Ecology Letters. **2011**. 14(8): 804-815.
- 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.
- Li, Y.; Schichtel, B. A.; Walker, J. T.; Schwede, D. B.; Chen, X.; Lehmann, C. M. B.; Puchalski, M. A.; Gay, D. A.; Collett Jr., J. L. Increasing importance of deposition of reduced nitrogen in the United States. PNAS. **2016**. 113 (21): 5874-5879.
- Michalak, A. M.; Anderson, E. J.; Beletsky, D.; Boland, S.; Bosch, N.S.; Bridgeman, T. B.; Chaffin, J. D.; Cho, K.; Confesor, R.; Daloğlu, I.; DePinto, J. V.; Evans, M. A.; Fahnenstiel, G. L.; He, L.; Ho, J. C.; Jenkins, L.; Johengen, T. H.; Kuo, K. C.; LaPorte, E.; Liu, X.; McWilliams, M. R.; Moore, M. R.; Posselt, D. J.; Richards, R. P.; Scavia, D.; Steiner, A. L.; Verhamme, E.; Wright, D. M.; Zagorski, M. A. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. PNAS. **2013**. 110 (16): 6448-6452.
- Newell, S.; Davis, T. W.; Johengen, T. H.; Gossiaux, D.; Burtner, A.; Palladino, D.; McCarthy, M. J. Reduced forms of nitrogen are a driver of non-nitrogen-fixing harmful cyanobacterial blooms and toxicity in Lake Erie. Harmful Algae. **2019**. 81: 86-93.
- Paerl, H. W.; Otten, T. G.; Kudela, R. Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum. Environmental Science & Technology. **2018**. 52 (10): 5519-5529.
- The Quality of Our Nation's Waters -- Nutrients and Pesticides: U.S. Geological Survey Circular 1225, 82 p. U.S. Geological Survey Circular, **1999**.
- Water Science School. Evapotranspiration and the Water Cycle. United States Geological Survey. **2019**.

University of Connecticut
School of Engineering

Acknowledgements

- Special thanks to Dr. Ellen Cooter (EPA retiree) who was responsible for conducting the CMAQ-Bidi and EPIC simulations; her contribution to this work has been immeasurable.

- 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.

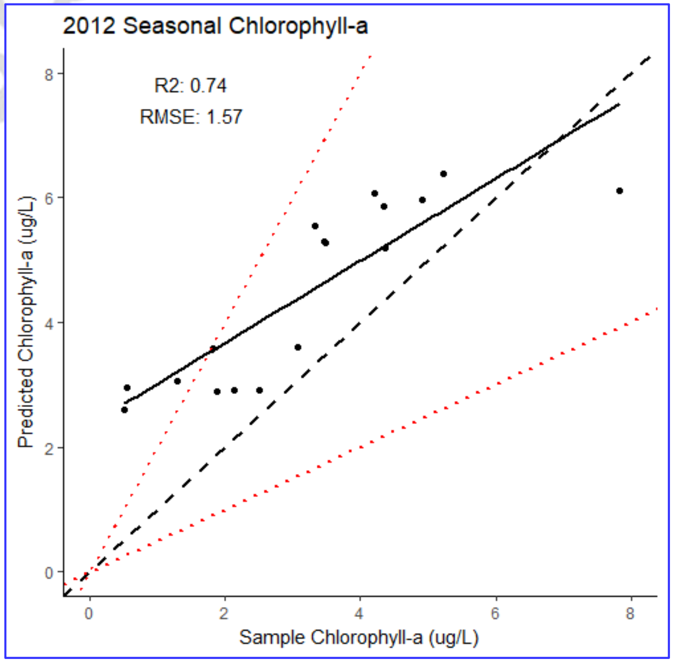
- ***Disclaimer:*** The views expressed in this presentation are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Questions:
Email: christina.feng_chang@uconn.edu

University of Connecticut
School of Engineering

ADDITIONAL MATERIAL

Discussion: Testing Approach



Separate RF model using:

- 2002-2011 to train and validate
- 2012 for testing

Chlor-α > 5 μg/L		**OBSERVATIONS**	
		YES	NO
MODEL	YES	2	7
	NO	0	8

Total points = 17
PC = 58.8%
POD_1 (Chlor-α >5 μg/L) = 100%
POD_2 (Chlor-α ≤5 μg/L) = 53.3%

- This test indicates **generalizability through time**
- Eutrophic conditions are identified 100% of the time
- Over 70% of variance in 2012 chlor-α measurements is explained by the RF model
- 82.4% of the model's predictions are within a factor of 2 of the obs
- Eutrophic vs. non-eutrophic conditions are correctly detected 58.8%

University of Connecticut
School of Engineering