PROJECT TITLE: Predicting Water Quality Extremes using Season-ahead Climate Information

INVESTIGATORS:

Paul Block (PI), Assistant Professor, University of Wisconsin – Madison; (608) 263-8792; paul.block@wisc.edu

Arun Kumar (co-PI), Chief, Development Branch, NOAA (NWS-CPC), College Park, Maryland; (301) 763-8000 x7579; Arun.Kumar@noaa.gov

NOAA GRANT NUMBER: NA14OAR4310270

PROJECT YEARS: 2014-2015

TIME PERIOD ADDRESSED BY REPORT: August 1, 2014 – April 30, 2015

I. Preliminary Materials

A. Research project objective

The project objective is to develop probabilistic season-ahead water quality outlooks by associating large-scale climate variable states and predictions with local in-land lake conditions, with particular emphasis on cyanobacteria levels, specifically extremes. We propose a forecast framework to characterize June-August seasonal cyanobacteria biomass variations and the inference of spatial and temporal effects on beach activities and closings. Using this framework, we hypothesize that skillful forecasts of seasonal water quality characterization are achievable to support recreational planning and promote public awareness and safety.

B. Stakeholders and decision makers

Project collaborators include

- Kirsti Sorsa at Public Health Madison and Dane County (Wisconsin)
- Emelia Wollenburg at Wisconsin Department of Health Service
- Clean Lakes Alliance, Madison, Wisconsin

Each stakeholder has a vested interest in cyanobacteria levels in the lake, lake health in general, and beach safety in specific.

C. Approach

The project approach includes the following steps:

1) Apply climate, hydrologic, and lake data to generate skillful water quality and beach closing forecasts at relevant time scales
2) Assess the potential value of using the forecasts and identify residual risk
3) Provide forecast information to stakeholders in an actionable form

Correlation between season-ahead (March – May) land-ocean-atmosphere variables and water quality parameters/beach closings will be performed to identify potentially skillful predictors for use in the statistical forecast models. Initially, water quality parameters will include cyanobacteria biomass, microcystis, and microcystin. The interactive plotting and analysis tool on the NOAA Physical Sciences Division's website will be used to establish regions of high correlation that are physically meaningful in influencing water quality/beach closings fluctuations in Lake Mendota. Correlations with common indices (e.g. North Atlantic Oscillation, Atlantic Tripole) will also be evaluated. A unique set of potential predictors for each water quality parameter and streamflow is expected.

To compliment the aforementioned potential predictors, outputs from the NOAA Climate Forecast System version 2 (CFSv2) dynamical model will likewise be correlated with seasonal water quality parameters and beach closings. These outputs will include springtime *predictions* of June-August hydroclimatic variables or forcing fields (e.g. precipitation, soil moisture, winds, sea-surface temperatures) to be added to the overall pool of potential predictors. Forecast bias is

a known issue for CFSv2 predictions, however no bias correction will be performed at this stage; rather predictors will be corrected empirically through the statistical forecast model.

Numerous statistical prediction modeling approaches will be evaluated through hindcast assessments for their ability to skillfully predict water quality parameters and beach closings, including parametric (e.g. linear and logistic regression, principle component analysis) and nonparametric (e.g. local polynomial) techniques. The principle component analysis technique is appealing as it minimizes over-fitting and multi-collinearities of predictors by nature of its construction. Particular attention will be paid to tuning the statistical model toward successfully predicting extremes. Models often produce outputs with low variance if conditioned solely on minimizing errors, thereby under-estimating extremes. Alternatives, such as weighting typical conditions that produce extremes, will be thoroughly evaluated. The model will produce an ensemble forecast for each summer season in the hindcast and future summer seasons, partially capturing system uncertainty, and allowing for a probabilistic interpretation and risk assessment based on water quality and beach closing forecast distributions. Model performance will be based on cross-validation and use common performance metrics. A physically-based lake model forced with hydro-meteorological predictions capable of producing expected water quality parameters will also be evaluated to compare with the statistically-based approach. (Colleagues in UW's Center for Limnology will share their expertise and model: General Lake Model with Framework for Aquatic and Biogeochemical and Aquatic Ecodynamic Modules.)

A half-day introduction and training session on the water quality forecast framework will be offered for local decision-makers and other interested organizations who have a stake in Lake Mendota water quality. This session will also be instrumental in defining future phases of this work. If supported by local institutions (e.g. Public Health Madison and Dane County, Clean Lakes Alliance, etc.) the categorical forecast, explanations, and caveats will be made available through their websites for public accessibility. We have allotted funds to support these technical aspects. Additionally, the predictions will be available through the Principle Investigator's UW-Madison website.

Overall project evaluation will be based upon successful completion of the following targets:

1. A viable and informative water quality forecast framework acceptable to local decision-makers and transferable to other communities
2. Forecast products and documentation made available to public through online access
3. Attracting numerous stakeholders (beyond project partners) to the half-day framework introduction and training session.
4. Final project report submitted to NOAA Climate Program Office and partners.

D. Matching funds/activities

Caitlin Soley, the UW-Madison graduate student working on this project, is a dual degree candidate in both Civil Engineering and the Water Resources Management Program (WRM) through UW-Madison's Nelson Institute. The WRM program has provided partial stipend and tuition support for Caitlin.

E. Partners

We are partnering with NOAA (NWS-CPC, Arun Kumar), the North Temperate Lakes Long Term Ecological Research Project (LTER, Emily Stanton; NSF-funded), and the Center for Limnology (Steve Carpenter) at UW-Madison. In next phases we will expand these partnerships to include the Department of Natural Resources (Gina LaLiberte), the United States Geological Survey (John Walker), the Global Lakes Ecological Observation Network (GLEON, Paul Hanson), and other partners at NOAA focused on HAB forecasts and ecological modeling (e.g. John Bratton, Richard Stumpf.) Additionally, project decision-makers are also serving as partners (Part I B above), providing data, guidance, feedback, etc.

II. ACCOMPLISHMENTS

A. Project timeline and tasks accomplished

Work in the first three quarters has focused on procuring and analyzing data, building predictive models, and partner engagement.

<u>Data Procurement</u>

Data has been obtained and synthesized from a number of sources, including:

- LTER: Lake Mendota water quality data, including in-lake phosphorus, cyanobacteria biomass and microcystis, secchi depth, temperature, phycocyanin, etc. Although these parameters have varying record lengths, Lake Mendota studies have been on-going for decades, providing a rich data source.

- NOAA: NCEP reanalysis data, including sea-surface temperature, sea-level pressure, geopotential height, air temperature and vector winds. Additionally, precipitation predictions from CFSv2 and calculated soil moisture.

- USGS: discharge data and phosphorus concentrations into Lake Mendota

- Local: local hydro-meteorology, including precipitation, ice cover days, and air temperature

- Public Health of Madison and Dane County: beach closings data (date, reason) for all public beaches on Lake Mendota starting in 2005. Also beach-specific microcystin data for the last 5-6 years.

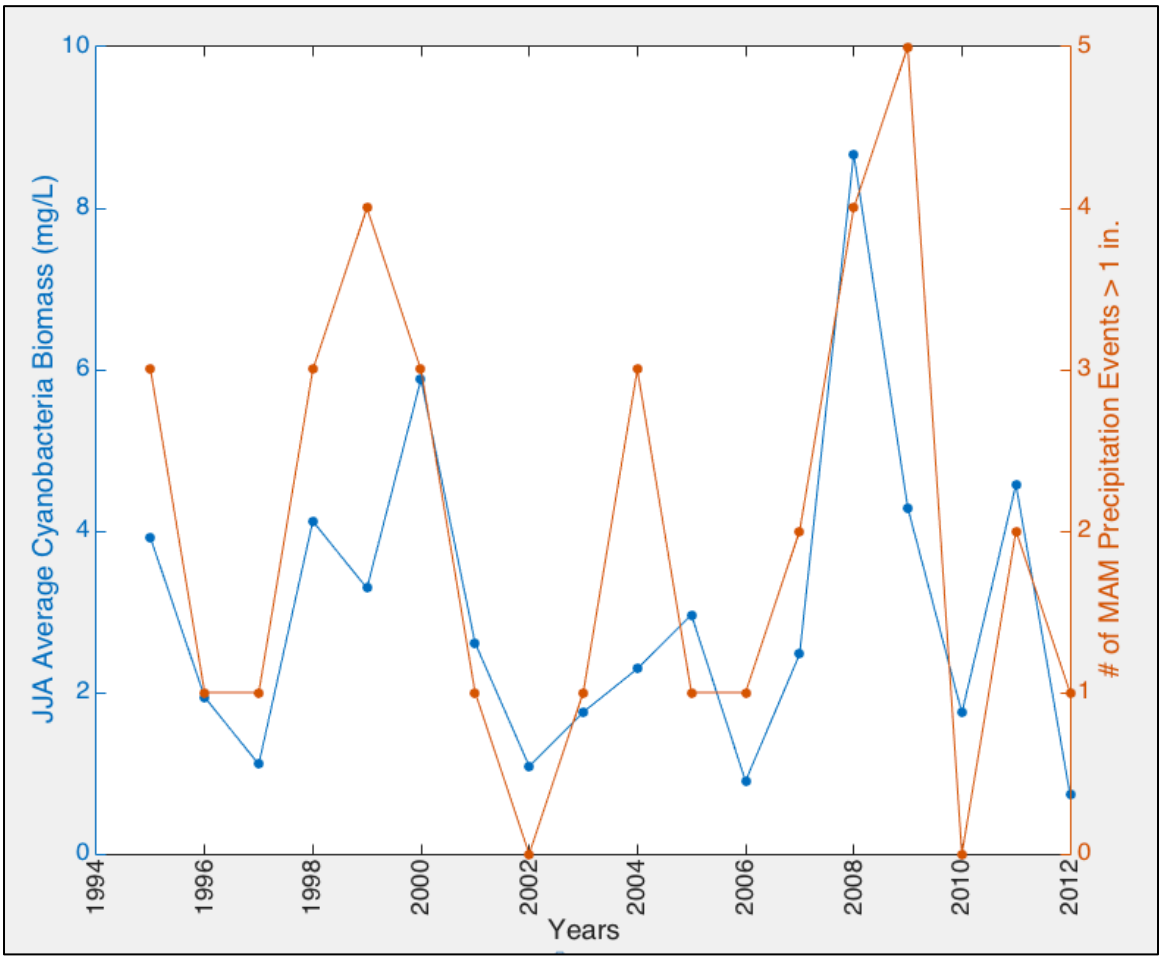
- Clean Lakes Alliance: citizen monitoring data (turbidity, ph, visual inspection, etc.) along Lake Mendota for the last two years.

Significant processing of data has been necessary to achieve a consistent format. (Requiring more time than initially planned.)

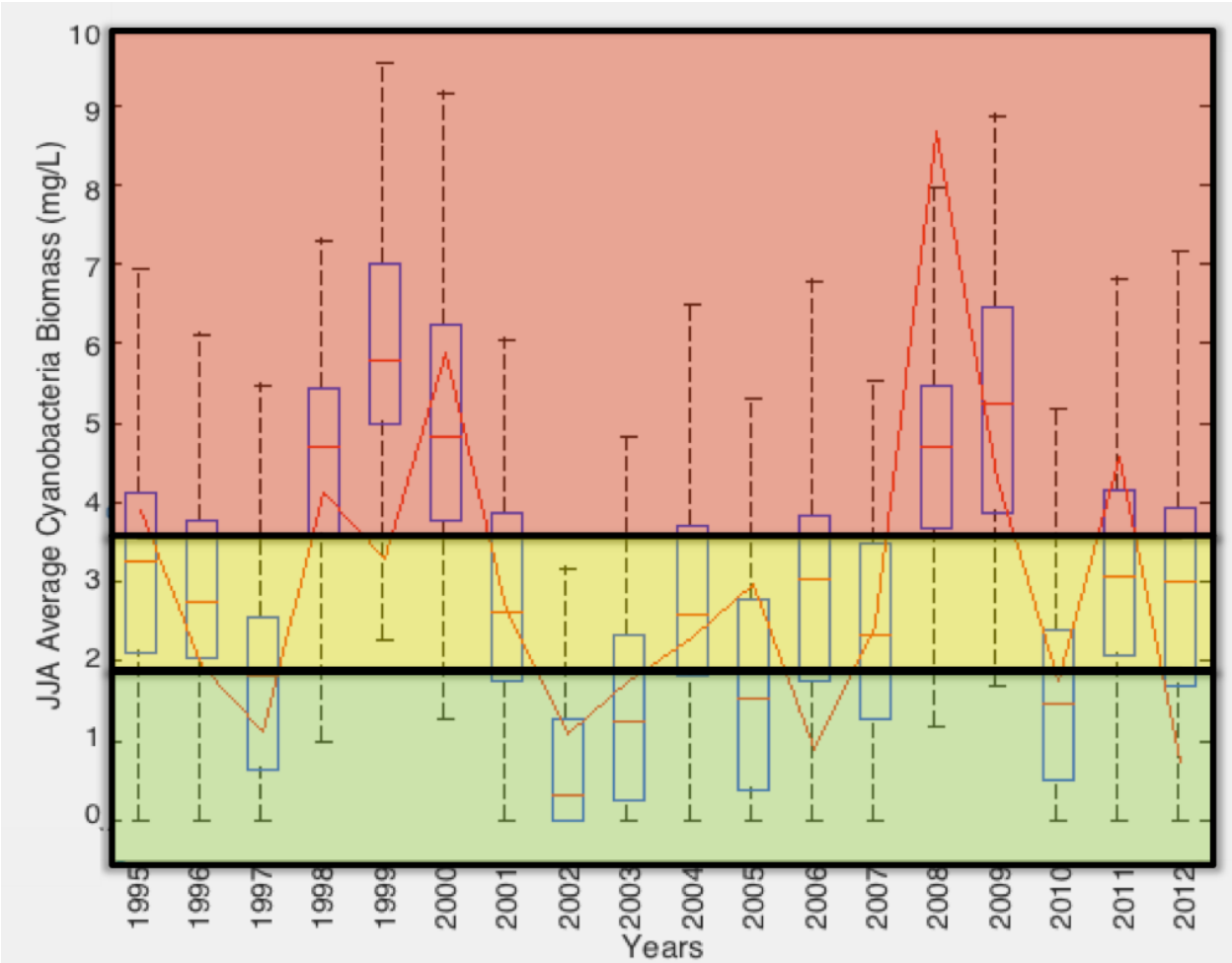
Numerous relationships have been evaluated based on the observational datasets. While it is clear that a number of factors influence cyanobacteria abundance in Lake Mendota (not all well

understood) a couple of unexpected aspects have surfaced. Phosphorus – hypothesized to be highly dominant in cyanobacteria production – does not strongly correlate. This includes for both in-lake (legacy) phosphorus and incoming phosphorus. Part of this weak relationship may be due to the spatial aspects of where data is observed. Incoming phosphorus is observed at the USGS gauging stations, while cyanobacteria is observed at the LTER buoy in the middle of the lake. Both are point measurements for parameters that have distinct spatial components. ENSO has also been found to be particularly well correlated with cyanobacteria levels in the lake (over the relatively short available time-series.) El Nino is typically associated with low cyanobacteria production and La Nina is typically associated with high production years. Longer time-series are necessary to confirm this but the available record certainly provides this indication.

A suite of variables across the March – May season have been evaluated for potential prediction of June – August average cyanobacteria biomas and microcystis biomass, including from LTER, NOAA, USGS, and other local variables listed above. Large-scale climate variables have been correlated with the predictands using the NOAA PSD tools. Regions of high correlation (e.g. SST, SLP) are identified and evaluated as potential predictors. While many March-May variables exhibit some level of correlation, a few stand out, including SSTs in the equatorial Pacific, the number of local precipitation events > 1-inch, and average discharge into the lake.



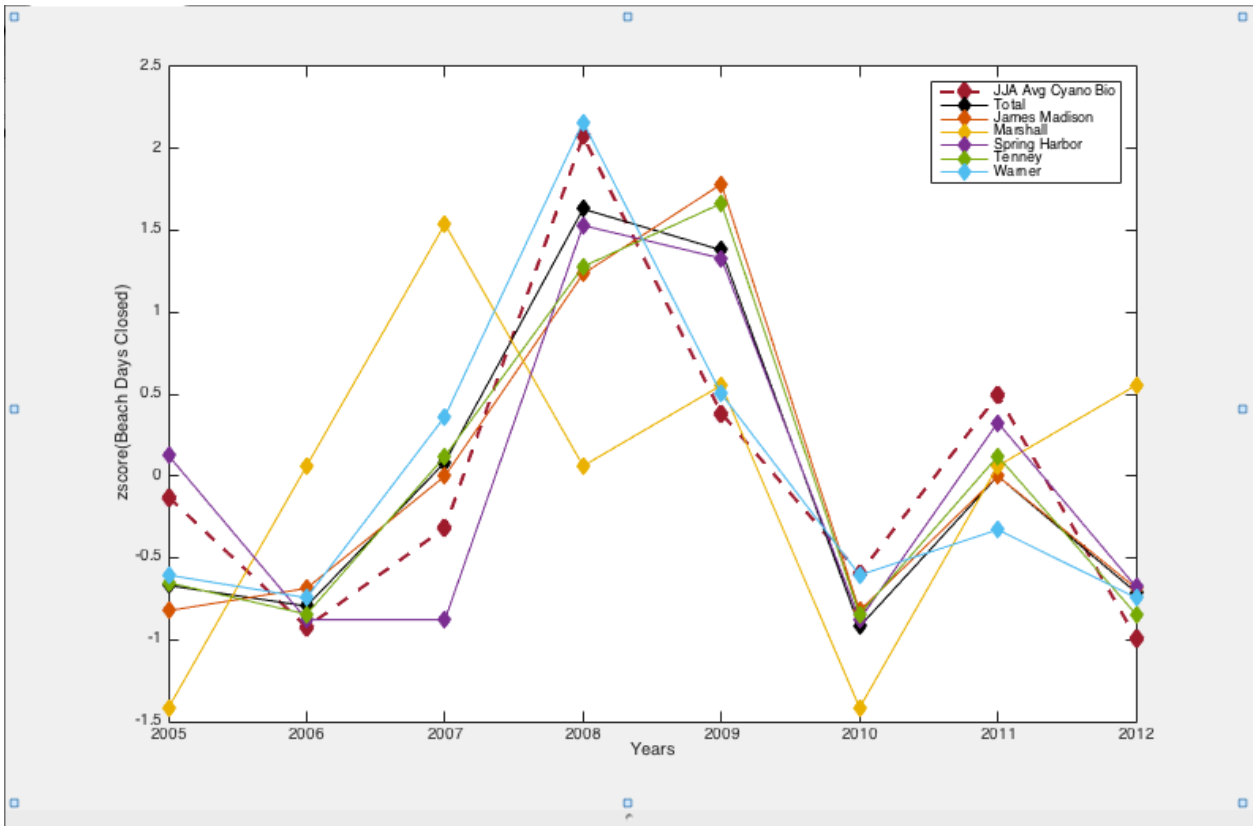
Both probabilistic logistic and principal component regression (statistical) model forms have been evaluated. The principal components approach simply using equatorial Pacific SSTs and local precipitation events to predict cyanobacteria biomass provides promising initial results, with correlation coefficients ~ 0.64 and rank probability skill scores (how well categorically – above, normal, below – the model preforms) ~ 0.66.



Similar findings are evident for microcystis biomass. Model development will continue, specifically by evaluating additional predictors and further explaining the physical relationship each predictors has with the predictand.

As previously discussed, the plausibility of predicting beach closings (both *events* and *number of days* across the season) are being explored. While prediction models are still in the early stage, there does appear to be a significant relationship between June-August average cyanobacteria abundance (at the LTER buoy) and the number of beach closing days. This is promising, given the heretofore skill in predicting cyanobacteria abundance. The beach closings present a spatial question as well. Initial evaluations tend to indicate eastern beaches are more positively correlated with cyanobacteria abundance than western beaches. Furthering these relationships

and predictive models – and eliciting feedback from stakeholders – will be an important remaining part of the project.



To compliment the statistical modeling approach, we have also explored using a physically-based lake model in a predictive mode to assess if prediction sill may be further augmented. This is also still in the early stages. The model has been previously developed and calibrated by colleagues at UW-Madison's Center for Limnology. The model will be driven by predicted discharge data from a USGS gauge.

B. Application of findings to inform decision-making

Application of findings to inform decision-making is primarily planned for the second year of the project. (We have requested a one-year no-cost extension.) However we have elicited and received feedback from stakeholders listed in Part I B above regarding approaches and identification of actionable prediction products. We have maintained regular correspondence with Kirsti Sorsa, including exchanges of preliminary research results.

C. Planned methods to transfer knowledge

In addition to informal meetings and email exchanges with project partners and stakeholders, which will continue throughout the project, we plan to hold a ½ day workshop formally introducing the forecast framework and providing training on application. This will be offered to

local decision-makers, partners, and other interested organizations having a stake in Lake Mendota water quality. This will also provide an additional opportunity for feedback on actionable prediction products – specifically form and format – which will greatly inform development of the online product.

D. Significant deviations from proposed work plan

The original version of this proposal was focused on Schuylkill River water quality, partnering with the Philadelphia Water Department. Given an institutional move by the PI to the University of Wisconsin – Madison, we submitted a proposed *revised work statement* to slightly refocus, changing the application site to Lake Mendota (Wisconsin) and updating project stakeholders and decision-makers. Project hypothesis and methodology are unchanged. This *revised work statement* was submitted prior to project initiation. Although the original work plan was scheduled to be completed in one year, we have requested a no-cost extension to allow more time for model development and stakeholder engagement. This is not a deviation, but rather a stretching of the project timeline.

E. Completed publications, white papers, or reports

There are no publications to date.

III. GRAPHICS: PLEASE INCLUDE THE FOLLOWING GRAPHICS AS SEPARATE ATTACHMENTS TO YOUR REPORT

A Powerpoint slide is included, depicting the overall framework for this project.

We presented this work at the American Water Resources Association (Wisconsin Chapter) annual meeting in March 2015. A pdf of the poster is attached.

Permission is hereby granted for use of any figures in this report.

Predicting Blooms: Can Cyanobacteria forecasts improve lake management?

Caitlin K. Soley and Professor Paul Block
Department of Water Resources Engineering, University of Wisconsin-Madison

Abstract

Cyanobacteria, also referred to as blue-green algae, are photosynthetic bacteria commonly polluting water bodies. Cyanobacteria are unique in their ability to produce toxic compounds called cyanotoxins, posing immediate health risks to the public and a rapid depletion of common nutrients including phosphorus, nitrates, and dissolved oxygen. This bacterium can render the lake insufficient to support a diverse ecological environment and pose serious threats to human health. Cyanobacteria has long been recorded as polluting Lake Mendota, prompting extensive research and interest. Presently, however, there is no long-term cyanobacteria abundance outlook that can provide governing bodies and users with advance warning of a potential threat in the upcoming season. Season-ahead cyanobacteria forecasts could allow for the advance of management techniques and improved public awareness. Two complimentary approaches are proposed to evaluate the the potential for informative cyanobacteria forecasts. A direct statistically based model will be developed utilizing probabilistic likelihood and severity of cyanobacteria for the June – August season. In addition, a dynamic modeling approach will also be considered; in this case, predicted climate and hydrologic conditions for the June - August season will be fed into a hydrodynamic-water quality model available through the University of Wisconsin-Madison limnology and microbial sciences departments.

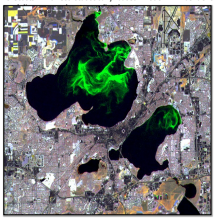
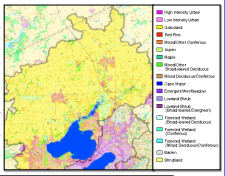


Fig. 1: Satellite photo of Lake Mendota algae bloom in 1999 produced by WisconsinView & UW Space Science and Engineering Center

Motivation

This project aims to create season-ahead forecasts that will allow various organizations and agencies to:

- Provide an early warning system to mangers to improve response efforts
- Further educate and inform users of potential health risks
- Mitigate and/or manage associated degradation of water quality



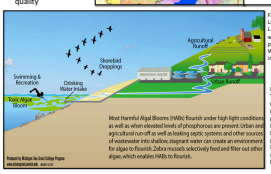


Fig. 2 (above): Landuse map for Lake Mendota watershed produced by Wisconsin DNR in 2005

Fig. 3 (bottom): The main factors influencing the growth of harmful algal blooms (HABs) produced by the Michigan Sea Grant College Program

Approach

Direct Statistical Approach

1. Synthesize observed data
2. Identify relevant season-ahead predictors from large-scale ocean-atmosphere states, dynamical model (general circulation models) predictions, remotely sensed data, and local land and water quality conditions
3. Construct statistical forecast utilizing hindcast assessments, linear regression, and principle component analysis.

Indirect Dynamical Approach

1. *Predict hydrologic variables via statistical forecast*
2. *Execute* hydrodynamic-water quality model (GLM FABM AED) using predicted variables

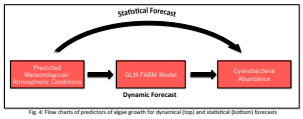


Fig. 4: Flow charts of predictors of algae growth for dynamical (top) and statistical (bottom) forecasts

Current Work

Statistical Forecasts:

Currently, season ahead forecasts have been developed for June-August (JJA) average Cyanobacteria biomass and July-August average Microcystis biomass with data provided by the North Temperate Lakes Long Term Ecological Research site on Lake Mendota. Both probabilistic models were developed using suites of local and global meteorological and atmospheric predictors (Fig 5). Fig 6a. & 6b. below illustrate the process used to developed each of the current forecasts. See Fig 7 & 8 for model results.

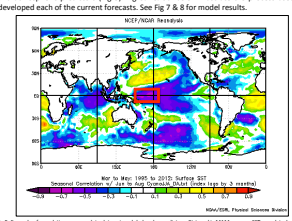


Fig 5: Example of correlation map used to determine global scale predictors. Pictured is MAM average SST correlated with JJA average Cyanobacteria biomass with a red box highlighting the general region chosen as the global predictor for the forecast

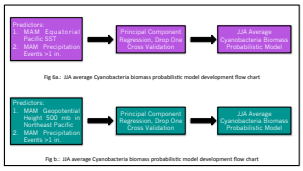


Fig 6a.: JJA average Cyanobacteria biomass probabilistic model development flow chart

Fig b.: JJA average Cyanobacteria biomass probabilistic model development flow chart

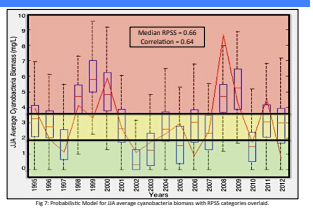


Fig 7: Probabilistic Model for JJA average cyanobacteria biomass with RPSS categories overlaid.

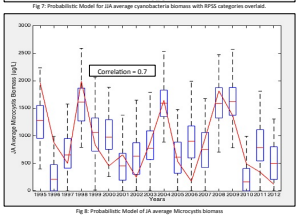


Fig 8: Probabilistic Model of JJA average Microcystis biomass

Dynamic Forecast:

The dynamic forecast will be developed by running a GLM FABM AED (General Lake Model coupled with Framework for Aquatic Biogeochemical and Aquatic Ecodynamics Models) calibrated through the University of Wisconsin - Madison Center for Limnology and Department of Microbial Sciences in a predictive outlook. Predicted meteorological inputs compiled from GCMs will be utilized as inputs to run the model as a forecast of Microcystis biomass. The statistical and dynamic forecast results for Ja average Microcystis biomass will be compared to asses which approach produced more skillful results.

Future Work

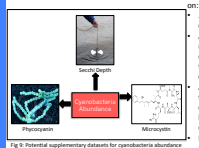


Fig 9: Potential supplementary datasets for cyanobacteria abundance

Future work of this project will focus on:

- Assess spatial variation of cyanobacteria in Lake Mendota
- Analyze supplementary datasets (Fig 9) to act as substitute for cyanobacteria abundance in forecasting to potentially increase confidence and skill
- Analyze supplementary datasets to provided additional annual detail (frequency and duration of high cyanobacteria abundance events)
- Determine optimal management communication framework for forecast results

Acknowledgements

- NOAA Award # NA14OAR4310270
- Kirsti Sorsa – Public Health Madison and Dane County