Probabilistic Forecasting of Harmful Algal Blooms in Western Lake Erie

Daniel R. Obenour[1,3], Andrew D. Gronewold[1,2], Craig A. Stow[2], Isabella Bertani[1], Cara E. Steger[1], Steven A. Ruberg[2], & Donald Scavia[1]
[1]University of Michigan; [2]NOAA Great Lakes Environmental Research Laboratory; [3]North Carolina State University

Sponsors: UM Water Center - Graham Sustainability Institute; Cooperative Institute for Limnology and Ecosystems Research; United States EPA - Great Lakes Restoration Initiative

Abstract

Over the past decade, there has been a dramatic rise in the magnitude of harmful algal blooms (HABs) in western Lake Erie. These cyanobacteria blooms have drawn attention to phosphorus loading, a common driver of freshwater productivity. However, it is unclear how much of the year-to-year variability in bloom size is explained by anthropogenic phosphorus loading, and how much variability is related to other factors, including weather/climate drivers and measurement error. Here, we aim to advance the state-of-the-art in HAB forecasting by explicitly quantifying uncertainties in late-summer bloom observations, and propagating them through a Bayesian modeling framework that relates bloom size to phosphorus load. Because of the need to accurately represent predictive uncertainty, different statistical formulations are critically evaluated through cross validation. A model based on a novel implementation of a gamma error distribution is found to provide the most realistic uncertainty characterization, as well as high predictive skill (Obenour et al., 2014). Our results also underscore the benefits of a hierarchical approach that allows us to assimilate data sets from multiple sources. Finally, our modeling analysis suggests that Lake Erie has become increasingly susceptible to large cyanobacteria blooms. We explore the nature of this change and assess potential biophysical explanations.

Bloom Response Data

Three sets of bloom observations are used in this study (Figure 1). The first set is developed from MERIS satellite remote sensing imagery, per a procedure developed by NOAA (Stumpf et al., 2012). The second set of observations is developed from a University of Toledo field monitoring program (Bridgeman et al., 2013). The third set is developed from SeaWiFS satellite remote sensing imagery by MTRI under contract to the Water Center, using procedures similar to Shuchman et al. (2006). Together, these sources cover the period of 1998-present.

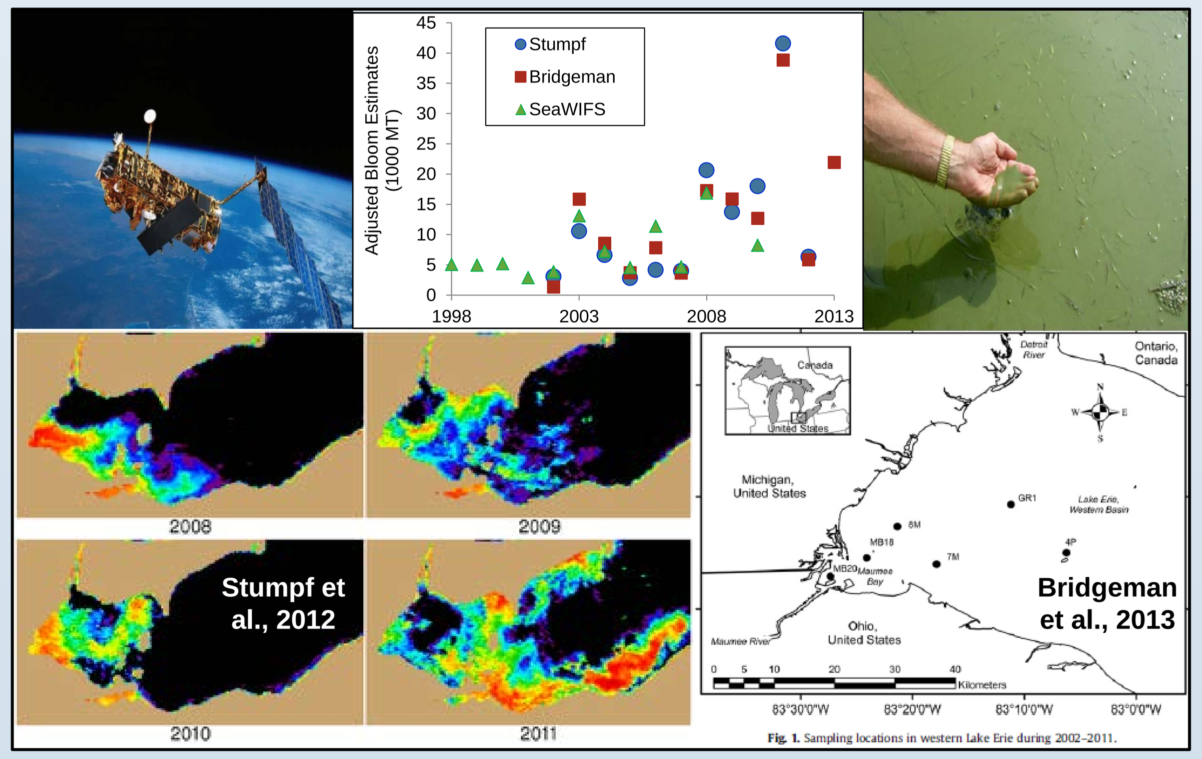

Figure 1: Illustration of MERIS remote sensing bloom estimates and field sampling bloom estimates.

Phosphorus Loading Data

Nutrient loads (TP and DRP) are generated from Maumee River nutrient concentration data collected by Heidelberg University's National Center for Water Quality Research (NCWQR, http://www.heidelberg.edu/academiclife/distinctive/ncwqr/data), and stream flow data collected by the United States Geological Survey (USGS, http://www.usgs.gov/water) at Waterville, Ohio (USGS Station 04193500). Nutrient concentration data are available on a near-daily basis, and missing concentrations are imputed using a local regression with flow.

Model Development

The model predicts annual peak bloom size in Western Lake Erie as a function of spring phosphorus loading from the Maumee River (Figures 2, 3). The model also includes a temporal term that captures changes in the lake`s susceptibility to HAB formation:

$$\hat{z}_i = \begin{cases} \beta_b + \beta_0 + \beta_w W_i + \beta_t T_i & \text{for} \beta_0 + \beta_w W_i + \beta_t T_i > 0 \\ \beta_b & \text{for} \beta_0 + \beta_w W_i + \beta_t T_i < 0 \end{cases}$$

where β_b, β_0, β_w, and β_t are model parameters that predict bloom size, $\hat{z}_i$, in year *i*, in terms of the bioavailable fraction of total phosphorus (TP) load, W_i, and model year, T_i. The parameter β_b is a background bloom level representing the bloom size in years of minimal phosphorus loading. The parameter β_0 is an intercept term (1000 MT bloom dry weight), and β_t represents how that intercept changes over time. The parameter β_w represents the unit increase in bloom size per unit increase in P load (1000 MT/mo).

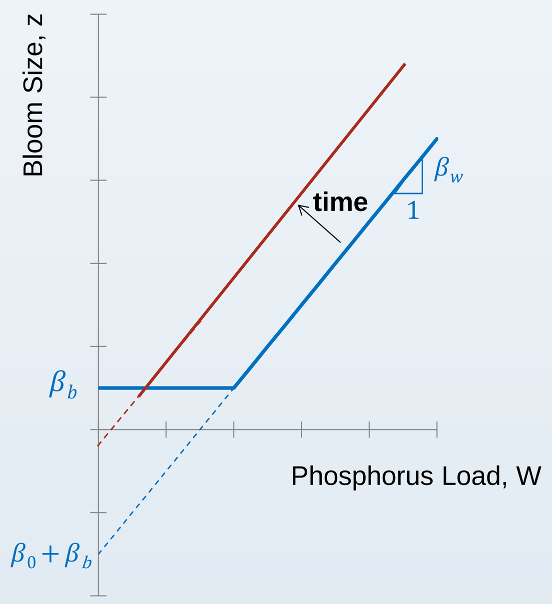

Figure 2: Deterministic load-bloom relationship

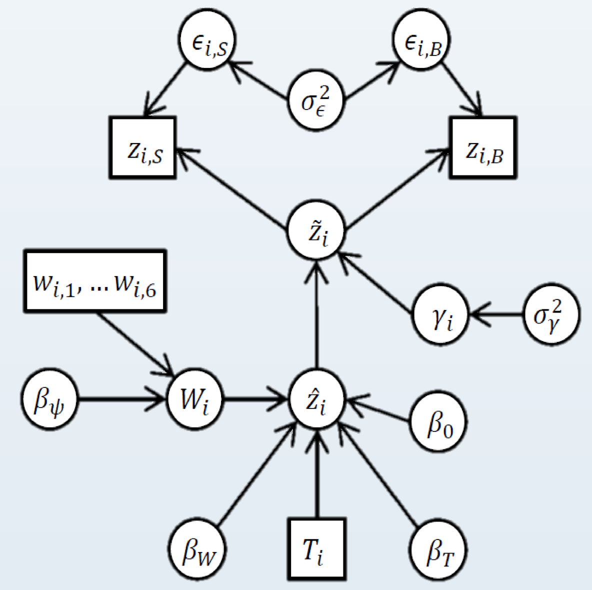

Figure 3: DAG representation of bloom forecast model. Circles represent the modeled variables and parameters, rectangles represent the input data, and arrows represent the conditional dependencies. Subscripts S and B refer to the Stumpf and Bridgeman observations (Obenour et al., 2014).

For each year, spring P load is determined as a weighted average of January to June (*m* = 1 to 6) monthly loads, based on the following equations, where β_ψ is a weighting parameter:

$$W_i = \frac{1}{\sum \psi_m} \sum_{m=1}^{6} w_{i,m} \psi_m \quad \text{where} \quad \psi_m = \begin{cases} 0 & \text{for } m \leq (\beta_\psi - 1) \\ m + 1 - \beta_\psi & \text{for } (\beta_\psi - 1) < m < \beta_\psi \\ 1 & \text{for } m \geq \beta_\psi \end{cases}$$

The bioavailable fraction of the TP load was estimated as the sum of the bioavailable fractions of Dissolved Reactive Phosphorus (DRP) and Particulate Phosphorus (PP) loads: Bioavailable P = DRP + θ*PP. DRP is expected to be 100% readily available to algae (Lambert, 2012), while only a fraction θ of the PP load is expected to be available (DePinto et al., 1981; Baker et al., 2014). The parameter θ was estimated probabilistically, together with other model parameters, through Bayesian inference (Obenour et al., 2014).

Two probabilistic expressions relate predicted values to the bloom observations, $z_{i,j}$:

$$z_{i,j} \sim Gamma[(\hat{z}_i + \gamma_i)^2/\sigma_\epsilon^2,\ (\hat{z}_i + \gamma_i)/\sigma_\epsilon^2]$$

$$\gamma_i \sim Gamma\left(\hat{z}_i^{\,2}/\sigma_\gamma^2, \hat{z}_i/\sigma_\gamma^2\right) - \hat{z}_i$$

The gamma distributions have shape (g_α) and rate (g_β) parameters (i.e., *Gamma*(g_α, g_β)) such that the mean and variance are g_α/g_β and g_α/g_β^2, respectively (Figure 4). Model prediction errors (γ_i) are drawn from a gamma distribution with variance σ_γ^2, and observation measurement errors are drawn from a gamma distribution with variance σ_ϵ^2. Here, subscript *j* differentiates between multiple observations of the same bloom.

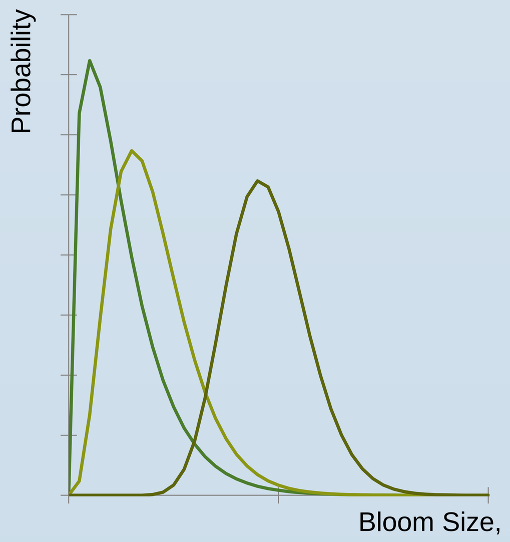

Figure 4: Example gamma probability distributions for different bloom sizes.

Results

Using a Bayesian calibration process, prior information about model parameters is updated based on model relationships and observed bloom data (Figure 5). The updated model estimate for θ suggests that around 50% of Maumee PP loading becomes bioavailable, though there is considerable uncertainty. Also note that trend parameter, β_t, is significantly positive. The load weighting parameter, β_ψ, suggests that loads occurring prior to February (<2) will not contribute directly to the summer bloom.

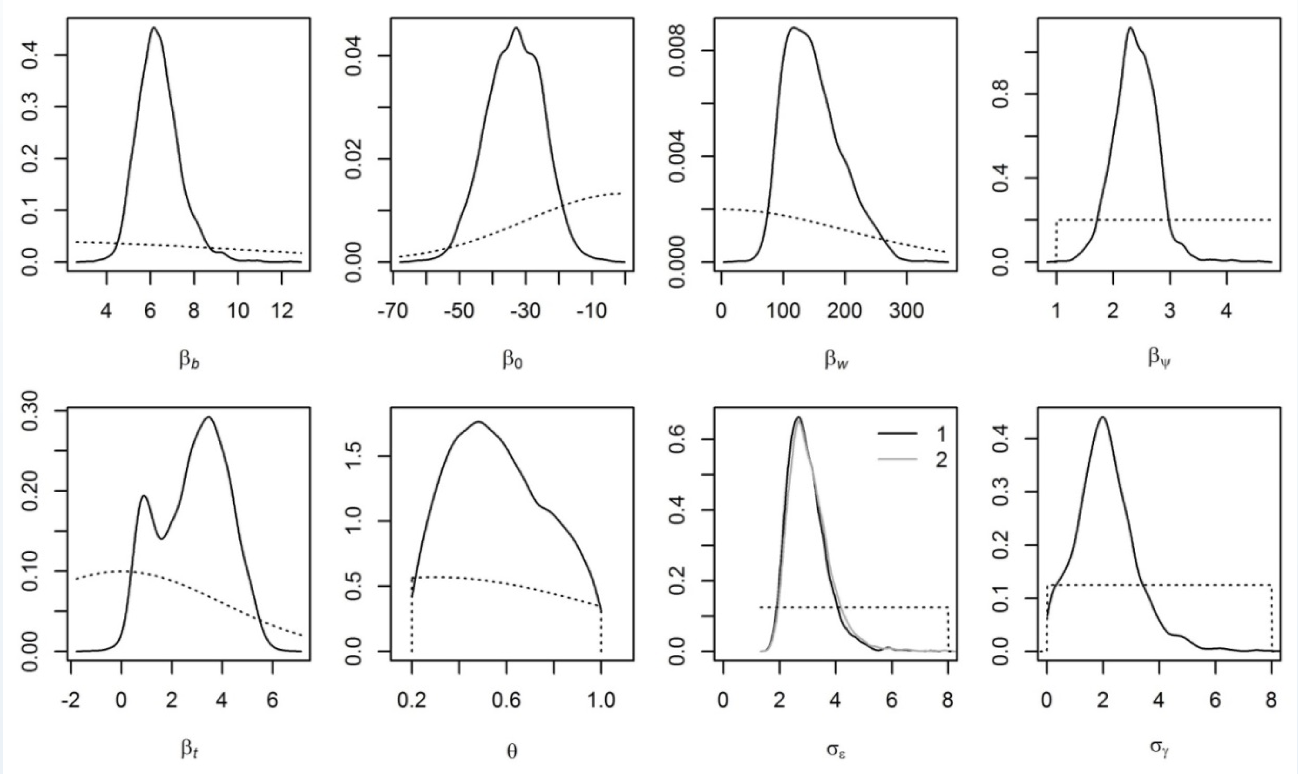

Figure 5: Prior (dashed) and marginal posterior (solid) model parameter distributions. The two distributions for σ_ϵ represent the measurement error for (1) Stumpf and Bridgeman observations and (2) SeaWiFS observations.

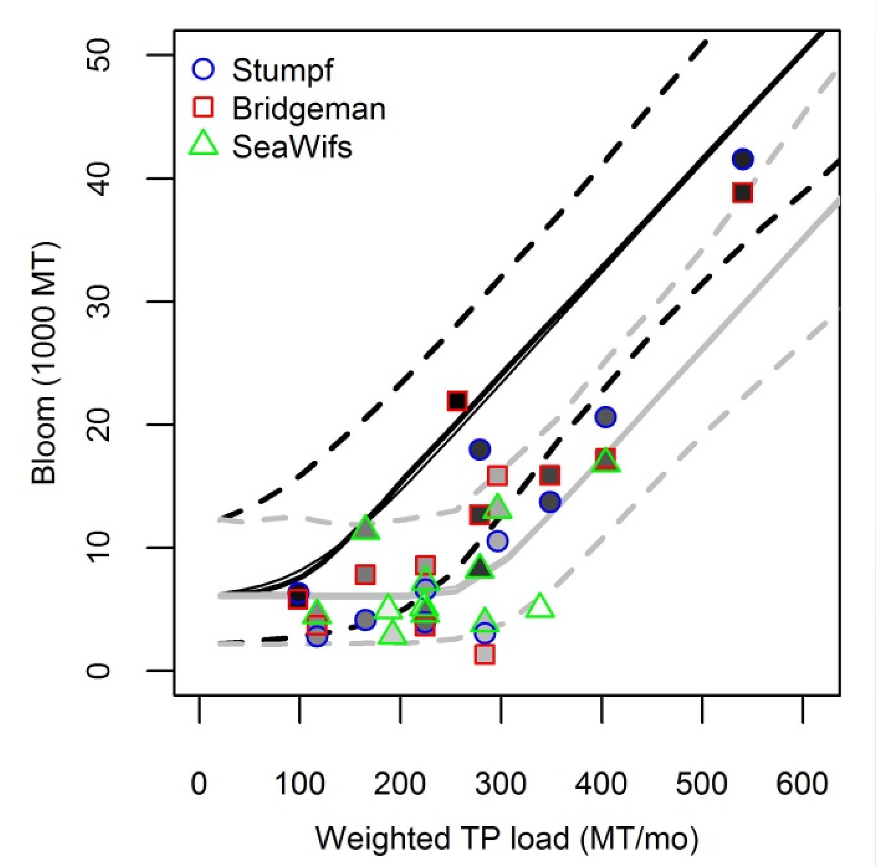

Figure 6: Load-response curve, with median prediction (thick lines), mean prediction (thin lines), and 95% predictive intervals (dashed). Grey lines: 2008 lake conditions, black lines: 2013 lake conditions. Bloom observations are shaded on a linear gradient from white (1998) to black (2013).

The model explains 91% of the annual variability in the Stumpf and Bridgeman bloom observations. To assess the model`s performance when predicting data not included in the calibration process, a leave-one-year-out cross-validation was carried out, and performance was similar (R^2=81%). Performance was somewhat lower for SeaWiFS observations, which are expected to have larger measurement error. The model is used to generate load-response curves (Figure 6) to aid in forecasting and management.

A change-point version of the model was also developed, but it does not explain long-term variability as well as the gradual (linear) trend. Nitrogen loading and temperature were also considered, but did not improve model performance. Future research will explore climate variability and invasive species as potential drivers of long-term bloom susceptibility.

Summary

- Bayesian hierarchical modeling provides a statistically rigorous way to accommodate multiple sets of bloom observations and non-linear model interactions. Furthermore, the use of gamma error distributions outperformed models based on the more-common normal and log-normal distributions (not shown).
- Spring TP load (February-June) is a significant predictor of late summer (August-September) peak bloom size, and results suggest that much of the particulate phosphorus (i.e., non-DRP) load may become bioavailable over the course of a summer.
- Results suggest Lake Erie has become increasingly susceptible to cyanobacteria blooms over the last decade. While neither water temperature nor nutrient loading (P or N) directly explained this trend, more research into long-term bloom dynamics would beneficial. The increasing trend in bloom susceptibility will likely complicate effort to reduce HABs in western Lake Erie.

References

(1) Baker, D. B., Confesor, R., Ewing, D. E., Johnson, L. T., Kramer, J. W., & Merryfield, B. J. (2014). Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: the importance of bioavailability. Journal of Great Lakes Research. doi: 10.1016/j.jglr.2014.05.001
(2) Bridgeman, T. B., Chaffin, J. D., & Filbrun, J. E. (2013). A novel method for tracking western Lake Erie Microcystis blooms, 2002–2011. Journal of Great Lakes Research, 39(1), 83-89.
(3) DePinto, J. V., Young, T. C., & Martin, S. C. (1981). Algal-available phosphorus in suspended sediments from lower Great Lakes tributaries. Journal of Great Lakes Research, 7(3), 311-325.
(4) Obenour, D. R., A. D. Gronewold, C. A. Stow, & D. Scavia (2014). Using a Bayesian hierarchical model with a gamma error distribution to improve Lake Erie cyanobacteria bloom forecasts, Water Resources Research, 50, 7847–7860.
(5) Shuchman, R., Korosov, A., Hatt, C., Pozdnyakov, D., Means, J., & Meadows, G. (2006). Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: a 7-year inter-annual analysis. Journal of Great Lakes Research, 32(2), 258-279.
(6) Stumpf, R. P., Wynne, T. T., Baker, D. B., & Fahnenstiel, G. L. (2012). Interannual variability of cyanobacterial blooms in Lake Erie. PLOS ONE, 7, e42444.

Acknowledgements: This project was funded in part by the Great Lakes Restoration Initiative (administered by the U.S. EPA) Synthesis, Observation, and Response (SOAR) project. We thank Richard Stumpf, Thomas Bridgeman, and Michigan Tech Research Institute (MTRI) for providing the bloom observation data. In addition, we thank Peter Richards and David Baker for assistance with the nutrient loading data.

AGU Poster Number: H11A-0848
Corresponding Author: Dan Obenour, drobenour@ncsu.edu