AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2014WR015616

Key Points:
- Bayesian model relates cyanobacteria bloom size to phosphorus load
- Algal bloom forecast uncertainty well characterized by gamma distribution
- Lake Erie increasingly susceptible to large cyanobacteria blooms

Correspondence to:
D. R. Obenour,
drobenour@ncsu.edu

Citation:
Obenour, D. R., A. D. Gronewold, C. A. Stow, and D. Scavia (2014), Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts, *Water Resour. Res.*, *50*, 7847–7860, doi:10.1002/2014WR015616.

Received 24 MAR 2014
Accepted 11 SEP 2014
Accepted article online 16 SEP 2014
Published online 8 OCT 2014

Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts

Daniel R. Obenour[1], Andrew D. Gronewold[2], Craig A. Stow[2], and Donald Scavia[3]

[1]Water Center, University of Michigan, Ann Arbor, Michigan, USA, [2]NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, USA, [3]Graham Sustainability Institute, University of Michigan, Ann Arbor, Michigan, USA

Abstract The last decade has seen a dramatic increase in the size of western Lake Erie cyanobacteria blooms, renewing concerns over phosphorus loading, a common driver of freshwater productivity. However, there is considerable uncertainty in the phosphorus load-bloom relationship, because of other biophysical factors that influence bloom size, and because the observed bloom size is not necessarily the true bloom size, owing to measurement error. In this study, we address these uncertainties by relating late-summer bloom observations to spring phosphorus load within a Bayesian modeling framework. This flexible framework allows us to evaluate three different forms of the load-bloom relationship, each with a particular combination of statistical error distribution and response transformation. We find that a novel implementation of a gamma error distribution, along with an untransformed response, results in a model with relatively high predictive skill and realistic uncertainty characterization, when compared to models based on more common statistical formulations. Our results also underscore the benefits of a hierarchical approach that enables assimilation of multiple sets of bloom observations within the calibration processes, allowing for more thorough uncertainty quantification and explicit differentiation between measurement and model error. Finally, in addition to phosphorus loading, the model includes a temporal trend component indicating that Lake Erie has become increasingly susceptible to large cyanobacteria blooms over the study period (2002–2013). Results suggest that current phosphorus loading targets will be insufficient for reducing the intensity of cyanobacteria blooms to desired levels, so long as the lake remains in a heightened state of bloom susceptibility.

1. Introduction

In response to excessive algal blooms, degraded water quality, and reduced hypolimnetic oxygen levels (hypoxia) in Lake Erie in the 1960s and 1970s [*Davis*, 1964; *Mortimer*, 1987], the United States (U.S.) and Canada initiated programs under the Great Lakes Water Quality Agreement [*International Joint Community (IJC)*, 1978] to lower point and nonpoint source loads of total phosphorus (TP), the limiting nutrient. Management actions implemented through these programs contributed to reduced loads [*Dolan*, 1993], decreased water-column TP concentrations [*DePinto et al.*, 1986; *Ludsin et al.*, 2001], smaller hypoxic areas [*Makarewicz and Bertram*, 1991; *Bertram*, 1993; *Charlton et al.*, 1993], and reduced phytoplankton biomass, especially cyanobacteria [*Makarewicz et al.*, 1989; *Makarewicz*, 1993]. Since the mid-1990s, however, these trends appear to have reversed [*Scavia et al.*, 2014] with significant increases, for example, in hypoxia [*Burns et al.*, 2005; *Hawley et al.*, 2006; *Zhou et al.*, 2013] and cyanobacteria blooms [*Conroy et al.*, 2008; *Bridgeman et al.*, 2012; *Stumpf et al.*, 2012]. In fact, in 2011, the lake-wide cyanobacteria bloom was the largest in recorded history, with a peak intensity over three times greater than any previously observed bloom [*Stumpf et al.*, 2012; *Michalak et al.*, 2013].

Appropriate identification and quantification of the drivers behind the shifting trend in Lake Erie water quality has important implications for water resources management, and has been a continued area of limnological and model forecast research. In a deconstruction of the 2011 bloom, for example, *Michalak et al.* [2013] showed that agricultural practices coupled with intense spring storms led to record-breaking P loads, and an extended period of weak lake circulation and warm quiescent conditions led to the development and persistence of the extensive bloom. Through analysis of recent and projected climate scenarios, they further suggested that all of these factors are consistent with expected future conditions, and without policy intervention to mitigate these impacts, similar blooms are likely to recur more frequently (for additional reading on Lake Erie's changing hydrologic condition, see *Gronewold and Stow* [2014]).

The Great Lakes Water Quality Agreement of 2012 included a call for minimizing the extent of hypoxic zones in Lake Erie associated with excessive phosphorus loading and maintaining cyanobacteria biomass at levels that do not produce concentrations of toxins that pose a threat to human or ecosystem health. As a key strategy, the agreement calls for review and update of associated phosphorus loading targets. *Rucinski et al.* [2014] developed response curves and loading targets for hypoxia. *Scavia et al.* [2014] compared those targets with others developed for cyanobacteria blooms [*Ohio EPA*, 2013] and suggested that while the latter target may help prevent cyanobacteria blooms, it may not be sufficient for reducing hypoxia significantly.

The cyanobacteria-related loading targets were based on exponential relationships between bloom size and spring TP load from the Maumee River [*Ohio EPA*, 2013; *IJC*, 2014], that were adapted from a similar exponential relationship between bloom size and spring Maumee River flow [*Stumpf et al.*, 2012]. While these relationships capture the general positive correlation between the log-transformed bloom size and spring TP load (or flow), they do not address the uncertainty in these relationships explicitly. However, that uncertainty is expected to be substantial due to multiple other factors affecting the interannual variability in observed bloom size, including summer wind patterns, temperature, nitrogen colimitation, and cyanobacteria bloom measurement error [e.g., *Michalak et al.*, 2013; *Chaffin et al.*, 2013; *Bridgeman et al.*, 2013].

Because there are currently only 12 years (2002–2013) of quantitative cyanobacteria bloom estimates for western Lake Erie, attempting to explicitly represent all (or even a broad set of) factors affecting bloom size within an empirical model would likely result in overparameterization and poor predictive performance [*Beck*, 1987]. In such cases where historical data are limited (as with many water resources systems), a parsimonious approach with an explicit representation of uncertainty is more warranted. In this study, we continue to use phosphorus load as the primary bloom predictor, as it is the factor that can be most reasonably addressed through watershed management. However, we redevelop the load-bloom relationship within a statistical framework where uncertainty is represented quantitatively. Further, we consider three different formulations of this relationship (using normal, lognormal, and gamma error distributions), and evaluate them based on their predictive skill and how realistically they characterize predictive uncertainty.

Because of the complexity of the model formulations used in this study, conventional approaches for developing probabilistic model solutions are not adequate, and a more flexible Bayesian approach is used instead. The efficacy of a Bayesian approach has been demonstrated in previous studies where empirical and mechanistic models have been used to predict algal blooms and hypoxia [e.g., *Qian et al.*, 2010; *Malve et al.*, 2007; *Cha et al.*, 2014; *Obenour et al.*, 2014]. In addition, this study uses a hierarchical approach [*Gelman and Hill*, 2007] to simultaneously calibrate the model to two sets of bloom estimates, one derived from remote sensing imagery [*Stumpf et al.*, 2012] and the other from in situ phytoplankton tows [*Bridgeman et al.*, 2013]. Hierarchical modeling is an effective tool for assessing multiple sources of uncertainty [*Qian et al.*, 2010], and here it allows us to rigorously account for variability due to measurement error, prediction error, and parameter uncertainty [*Cressie et al.*, 2009]. The utility of hierarchical modeling has been demonstrated in studies modeling the spatiotemporal variability in climatological and hydrological data sets [e.g., *Cooley and Sain*, 2010; *Najafi and Moradkhani*, 2013]. While Bayesian hierarchical modeling is also an important area of water quality modeling research [e.g., *Gronewold and Borsuk*, 2010; *Robson*, 2014], this is the first harmful algal bloom study that we are aware of to demonstrate a gamma error distribution for representing predictive uncertainty, and a hierarchical approach for assimilating multiple observation data sets.

2. Materials and Methods

2.1. Bloom Observations

This study uses two sets of cyanobacteria bloom measurements, hereafter referred to as the "observations." The first set of observations was developed using satellite remote sensing [*Wynne et al.*, 2010; *Stumpf et al.*, 2012]. For each year, the observed bloom was determined for the 30 day period with the highest mean cyanobacteria level, reported in units of "cyanobacteria index" (CI), where one CI corresponds to approximately 1.2×10^{20} cyanobacteria cells [*Stumpf et al.*, 2012]. Because the average dry weight of Lake Erie *Microcystis* is approximately 40 pg/cell [see *Bridgeman et al.*, 2013, Figure 2], one CI is equivalent to approximately 4800 metric tons (MT) cyanobacteria dry weight. While these bloom estimates were developed for all of Lake Erie, the blooms were largely confined to the lake's western basin (nominally west of Pelee Island) in almost all years [see *Stumpf et al.*, 2012, Figure 3]. An exception was October 2011, when eastward currents moved

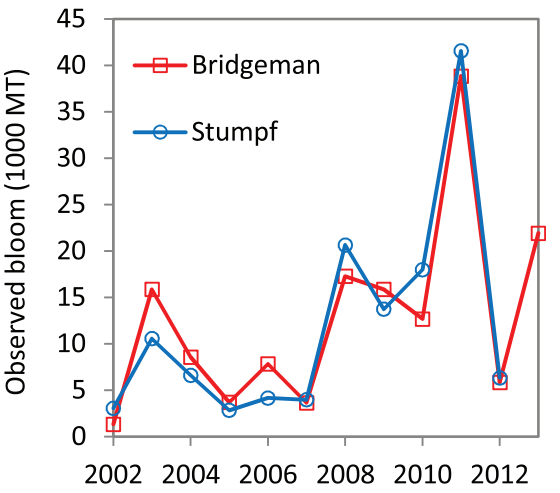

Figure 1. Annual western basin cyanobacteria bloom measurements developed by *Stumpf et al.* [2012] and *Bridgeman et al.* [2013], converted to units of mass (dry weight).

the bloom into the central basin where it continued to expand [*Michalak et al.*, 2013]. For consistency, this study uses the September 2011 bloom estimate (24 August to 22 September), when the bloom was largely confined to the western basin. This is also more temporally consistent, as the bloom peaked prior to October in all other study years.

The second set of bloom observations was developed from western basin phytoplankton tows [*Bridgeman et al.*, 2013]. Here the yearly bloom observations are reported in units of summer biovolume production (mL m^{-2}). Because there is considerable uncertainty in how to integrate these estimates over space and time, we treat them as measurements of relative bloom intensity, and scale them directly to the Stumpf estimates so that the means of the two data sets match, using a conversion factor of 1.1 MT/1.0 mL m^{-2}. Thus, while the scaled Bridgeman observations do not provide new information regarding absolute bloom size, they do provide an independent assessment of the year-to-year variability in relative bloom size.

Based on the conversions provided above, the two sets of annual bloom estimates can be compared over time (Figure 1). For brevity, the two sets are hereafter referred to as the "Stumpf" and "Bridgeman" observations. Whereas the Stumpf observations for 2002–2011 were obtained from *Stumpf et al.* [2012], the 2012 observation was provided directly from Stumpf (R. P. Stumpf, personal communication, 2014). Similarly, the Bridgeman observations for 2002 through 2011 were obtained from *Bridgeman et al.* [2013], and the 2012 and 2013 observations were provided directly from Bridgeman (T. B. Bridgeman, personal communication, 2013).

2.2. Predictor Variables

Nutrient loads are determined 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). In this study, NCWQR mean daily nutrient concentrations are multiplied by USGS mean daily flows to determine daily loads (MT/d). TP, dissolved reactive phosphorus (DRP), and nitrate loads were all considered as potential bloom predictors, but this manuscript focuses primarily on TP, which was found to be the most effective bloom predictor.

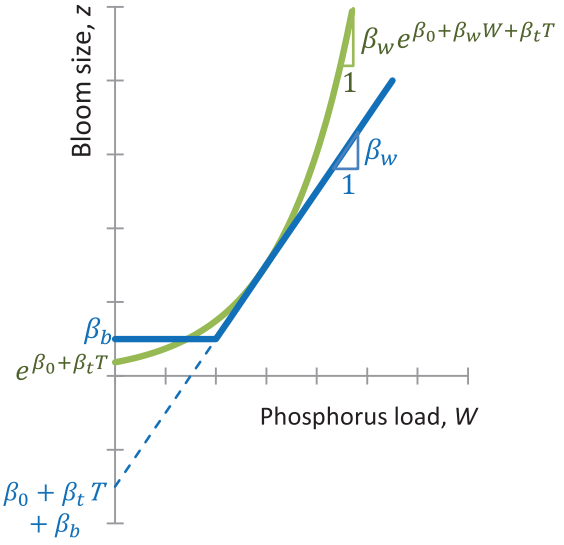

Figure 2. Deterministic relationship between bloom size and TP load under the normal and gamma model formulations (blue) and the log-normal model formulation (green).

TP concentration data were missing from approximately 6% of days throughout the study period. Each missing TP concentration was imputed using a linear regression with flow, calibrated from the closest 20 days with nonmissing data (where 20 days was determined optimal through cross validation, $R^2 = 65\%$). Because only a small

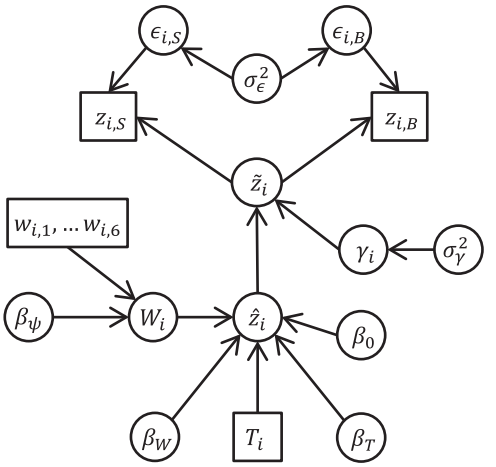

Figure 3. DAG representation of bloom forecast model. Circles represent the modeled variables and parameters, rectangles represent the input data, and arrows represent the conditional dependencies. All symbols are described in sections 2.3 and 2.4; subscripts *S* and *B* refer to the Stumpf and Bridgeman observations, respectively.

percentage of days had missing TP concentrations, the uncertainty associated with missing TP concentrations is considered negligible. DRP and nitrate were imputed similarly, but because these nutrient concentrations do not have a strong relationship with flow, the imputation was based simply on the mean of the closest 10 days with nonmissing data.

2.3. Candidate Models

Three candidate regression models are evaluated in this study. All of the models take the basic form:

$$z_{i,j} = \hat{z}_i + \gamma_i + \epsilon_{i,j} \tag{1}$$

where (for a given year *i* and observation set *j*) $z_{i,j}$ is a bloom observation, $\hat{z}_i$ is a deterministic bloom prediction based on nutrient loading, γ_i is a year-specific stochastic error term, and $\epsilon_{i,j}$ is an observation-specific stochastic error term. The distinction between year-specific errors and observation-specific errors is important because there are multiple bloom observations for each annual bloom, and the true bloom size ($\tilde{z}_i = \hat{z}_i + \gamma_i$) is unknown. From a hierarchical modeling perspective, γ_i is a yearly "random effect" [*Clark*, 2007] representing deterministic model error in predicting the true bloom size. Importantly, this approach addresses intraclass correlation [*Gelman et al*., 2004] that exists because the Stumpf and Bridgeman observations for a given year are not independent (i.e., they are both imperfect measurements of the same bloom, and are thus assigned the same predictive error component, γ_i).

The three candidate models are differentiated by their deterministic form (Figure 2) and by the probability distributions used to represent the stochastic error terms (γ and ϵ). These differences give rise to different predictions and predictive uncertainties (i.e., different widths of uncertainty bounds). The first bloom model, hereafter referred to as the "normal model," uses the common assumption of normally distributed measurement errors and random effects:

$$z_{i,j} \sim N\left(\hat{z}_i + \gamma_i, \sigma_\epsilon^2\right) \tag{2}$$

$$\hat{z}_i = \begin{cases} \beta_b + \beta_0 + \beta_w W_i + \beta_t T_i & \text{for } \beta_0 + \beta_w W_i + \beta_t T_i > 0 \\ \beta_b & \text{for } \beta_0 + \beta_w W_i + \beta_t T_i < 0 \end{cases} \tag{3}$$

$$\gamma_i \sim N\left(0, \sigma_\gamma^2\right) \tag{4}$$

where the deterministic prediction, $\hat{z}_i$, is a function of parameters β_b, β_0, β_w, and β_t and the weighted TP load, W_i (described in section 2.4). For years of relatively little nutrient loading, such that $\beta_0 + \beta_w W_i + \beta_t T_i < 0$, the bloom size is determined to be at background level, β_b, thereby preventing negative deterministic bloom predictions. The deterministic form of the model can be represented schematically (Figure 2). As shown, β_w defines the rate of change in bloom size per unit of phosphorus load (MT bloom)/(MT/mo), and $\beta_b + \beta_0 + \beta_t T$ is the *y* intercept that varies with time, *T*. Here $\beta_b + \beta_0$ is the intercept for 2007, and *T* varies from −5 to +6 (2002–2013); by making β_0 representative of the middle of the study period, we reduce parameter correlation and make the calibration process more computationally efficient. Finally, the parameter σ_ϵ is the standard deviation associated with measurement error, and the parameter σ_γ is the standard deviation associated with the yearly random effects (i.e., model prediction error). Model structure is illustrated in the form of a directed acrylic graph (DAG, Figure 3).

The second bloom model, hereafter referred to as the "gamma model," is identical to the normal model in deterministic form (Figure 2) and model structure (Figure 3), but it assumes measurement errors and random effects are distributed according to gamma distributions:

$$z_{i,j} \sim Gamma\left[(\hat{z}_i+\gamma_i)^2/\sigma_\epsilon^2,\quad (\hat{z}_i+\gamma_i)/\sigma_\epsilon^2\right] \tag{5}$$

$$\hat{z}_i = \begin{cases} \beta_b+\beta_0+\beta_w W_i+\beta_t T_i & \text{for } \beta_0+\beta_w W_i+\beta_t T_i > 0 \\ \beta_b & \text{for } \beta_0+\beta_w W_i+\beta_t T_i < 0 \end{cases} \tag{6}$$

$$\gamma_i \sim Gamma\left(\hat{z}_i{}^2/\sigma_\gamma^2,\ \hat{z}_i/\sigma_\gamma^2\right)-\hat{z}_i \tag{7}$$

where $z_{i,j}$ is now modeled as a gamma distribution with 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 $g_\alpha/g_\beta{}^2$, respectively. The random effects (equation (7)) are modeled as a gamma distributions centered at zero by subtracting $\hat{z}_i$. By zero centering the random effects, the gamma model formulation is more comparable to the normal model formulation (equations (2)–(4)), with the deterministic component (equation (6)) clearly distinguished from the random effect component.

The third bloom model, hereafter referred to as the "log-normal model," assumes measurement errors and random effects are normally distributed when predicting a log-transformed response:

$$\ln(z_{i,j}) \sim N\left(\hat{z}_{i;L}+\gamma_{i;L}, \sigma_{\epsilon;L}^2\right) \tag{8}$$

$$\hat{z}_{i;L} = \beta_0+\beta_w W_i+\beta_t T_i \tag{9}$$

$$\gamma_{i;L} \sim N\left(0, \sigma_{\gamma;L}^2\right) \tag{10}$$

where $\hat{z}_{i;L}$ is the deterministic bloom prediction on the log scale which can be back-transformed to the original scale using $\hat{z}_i = e^{\hat{z}_{i;L}}$ (as the median of the predictive distribution, Figure 2). Because the log transformation precludes negative predictions, there is no background bloom parameter β_b, as it would be indistinguishable from β_0. The parameters $\sigma_{\epsilon;L}$ and $\sigma_{\gamma;L}$ are the measurement error and random effect standard deviations, respectively, on the log scale. While model uncertainty is constant on the log scale, it is proportional to bloom size (i.e., multiplicative) on the original scale.

2.4. Weighted Phosphorus Load

The optimal loading period for predicting bloom size is also assessed probabilistically in this study. All of the candidate models use a weighted TP load, W_i, determined as follows:

$$W_i = \frac{1}{\sum \psi_m} \sum_{m=1}^{6} w_{i,m}\psi_m \tag{11}$$

$$\psi_m = \begin{cases} 0 & \text{for } m \le (\beta_\psi-1) \\ m+1-\beta_\psi & \text{for } (\beta_\psi-1) < m < \beta_\psi \\ 1 & \text{for } m \ge \beta_\psi \end{cases} \tag{12}$$

where $w_{i,m}$ is the TP load corresponding to month (m) and year (i), ψ_m is the weighting value for month m, as determined by β_ψ, the weighting parameter. The months of January to June ($m = 1$ to 6) are considered in the model, and β_ψ is thus constrained to the range [1, 6]. For example, if $\beta_\psi = 4.6$, then loads for January–March all receive weights of zero, April receives a weight of 0.4, and May and June both receive a weight of one. Thus, β_ψ can be thought of as a temporal threshold, prior to which loading does not contribute to the late-summer algal bloom. Although more sophisticated methods could be used for developing the monthly weights, this simple approach is advantageous for minimizing the number of model parameters. Adding later months (i.e., July–August) was considered, but this did not improve model performance; and avoiding later months allows for an earlier bloom forecast.

2.5. Bayesian Calibration

Model parameters were estimated using a Markov Chain Monte Carlo (MCMC) implementation of Bayes Theorem. Specifically, MCMC sampling was performed using WinBUGS software [*Lunn et al.*, 2000] called from R [*R Development Core Team*, 2008] via R2WinBUGS [*Sturtz et al.*, 2005; *Gelman and Hill*, 2007]. The MCMC sampling was performed in three parallel "chains" of up to 200,000 samples each, and the first half

Table 1. Prior Distributions for the Parameters of the Three Candidate Models[a]

Symbol	Description	Units	Normal Model	Gamma Model	Log-Normal Model
β_0	Bloom (relative) intercept	z	N(0,30)	N(0,30)	N(0,5)
β_w	Bloom production	z/(1000 MT/mo)	N(0,100)	N(0,100)	N(0,10)
β_b	Bloom baseline	z	N(0,10)	N(0,10)	n/a
β_t	Temporal trend	z/yr	N(0,4)	N(0,4)	N(0,1)
β_ψ	Load weighting		U(1,6)	U(1,6)	U(1,6)
σ_γ	Random effect standard deviation	z	U(0,8)	U(0,8)	U(0,2)
σ_ϵ	Measurement standard deviation	z	U(0,8)	U(0,8)	U(0,2)

[a] z is the bloom size with units of 1000 MT for the normal and gamma models and ln(1000 MT) for the log-normal model.

of each chain was removed as a "burn-in period" [*Gelman et al*., 2004]. The remaining chain portions were then thinned to 1000 samples each to reduce autocorrelation, and checked to ensure that they had converged on equivalent posterior parameter distributions. Convergence was evaluated using the $\hat{R}$ statistic, which is the square root of the ratio of a parameter's marginal posterior variance to its within-chain variance [*Brooks and Gelman*, 1998]. When $\hat{R}$ is close to one (i.e., $\hat{R} \leq 1.1$) for all model parameters, convergence is considered achieved.

The parameters for the three candidate models and their prior distributions are listed in Table 1. Here the prior distributions represent probabilistic expectations of parameter values based on a priori system knowledge. Because of the empirical nature of these models, vague zero-centered normal distributions, N(mean, standard deviation), are used for most parameters. For the normal and gamma models, β_b and β_w are constrained to be positive, because the model calibration becomes insensitive to these parameters if they are negative (and the solution is nonsensical), preventing posterior convergence. In addition, wide uniform distributions, U(lower bound, upper bound), are used for the error parameters σ_ϵ and σ_γ, as recommended by *Gelman* [2006]; and an informative uniform distribution is used for β_ψ to constrain it as described in section 2.4. Finally, sampling of the gamma distribution in equation (7) is censored such that $(\hat{z}_i + \gamma_i) > 750$ MT because the MCMC algorithm fails when attempting to sample from a gamma distribution with a very small mean (equation (5)). As the smallest bloom observation is well above 750 MT, this restriction does not adversely affect model performance.

2.6. Model Assessment and Comparison

Models are assessed based on their predictive skill and how accurately they characterize predictive uncertainty. Predictive skill is measured using the coefficient of determination (R^2), which is the percent of the variance in the observations that is resolved by the model predictions (i.e., the means of the predictive distributions) relative to a null (constant-only) model [*Gelman and Hill*, 2007]. We note that this is equivalent to the definition of R^2 proposed by *Nash and Sutcliffe* [1970], but it is different from the square of the Pearson's correlation coefficient (denoted r^2 in this text). R^2 is a common and intuitive measure of predictive skill, useful for making comparisons across models with different distributional assumptions, because it uses only the means of the predictive distributions. However, for this same reason, R^2 is not useful for assessing whether a model adequately characterizes predictive uncertainty. *Posterior predictive p values* can be used to assess the appropriateness of the overall posterior distribution [*Meng*, 1994; *Elmore*, 2005]. A predictive p value is determined for each observation based on the portion (i.e., cumulative probability) of the Bayesian posterior predictive distribution that exceeds the observed value, as illustrated graphically by *Gronewold et al*. [2009]. If the model accurately characterizes predictive uncertainty, then predictive p values should be uniformly distributed between zero and one. For this study, histograms of predictive p values are developed (using a bin size of 0.2) to visually assess their uniformity. However, because of the small number of observations available for this study ($n = 23$), some deviation from uniformity is expected. To gage the severity of this deviation, 93% confidence bands [*Härdle et al*., 2004] for bin frequency are developed through simulation-based approximation [*Westfall and Henning*, 2013]. Here 5000 histograms are generated from random uniformly distributed data, and the maximum and minimum frequency (i.e., bin count) of each histogram is recorded. The 93% confidence bands are then determined as the extrema values that completely contain all but 7% of the generated histograms. Thus, if a model properly characterizes uncertainty, its histogram of predictive p values has a 93% probability of falling completely within these bands. (The small n and discrete nature of histogram bin frequencies precludes determination of a 95% confidence band.) For

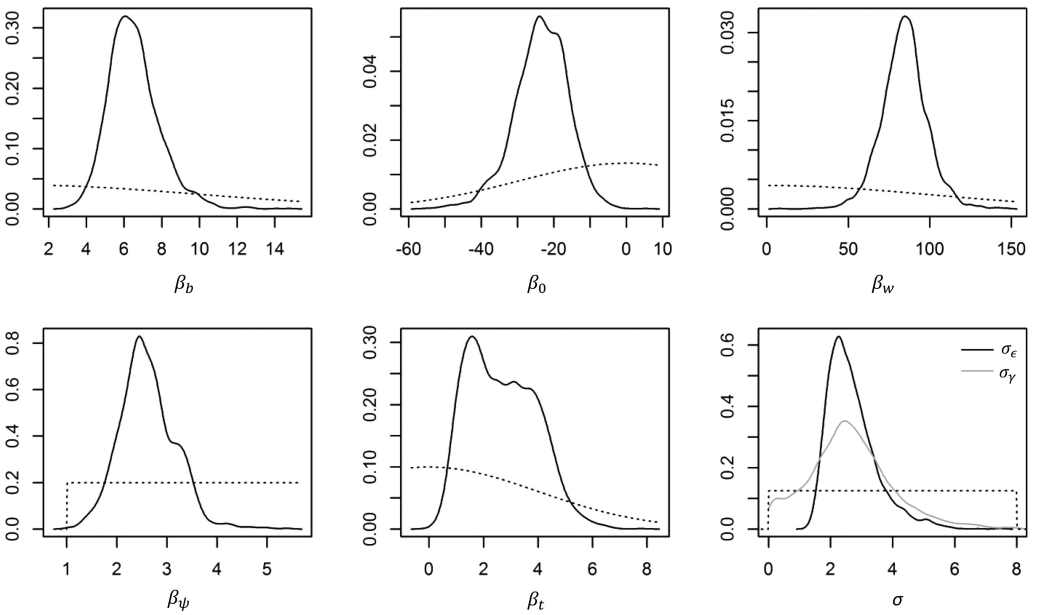

Figure 4. Prior (dashed) and marginal posterior (solid) parameter distributions for the gamma model (parameter units and descriptions same as Table 1).

reference, we also report the root mean square error (RMSE) and mean absolute error (MAE) of the different models. Generally, the "best" model is determined to be the model that has the highest predictive performance (considering R^2, RMSE, and MAE) while also providing realistic uncertainty quantification.

To test the robustness of each model, a leave-one-year-out cross validation (CV) is also performed. In CV, the observations for each year are predicted after removing those observations from the calibration data set and recalibrating the model to the reduced data set. The model's CV performance is a better measure of how well it will perform when predicting "out-of-sample" (i.e., future) conditions [*Elsner and Schmertmann*, 1994; *Chatfield*, 2006].

3. Results

3.1. Model Parameterization

The marginal posterior distribution for each model parameter can be illustrated graphically (Figure 4, for the gamma model only, for brevity) and summarized by its mean and 95% credible interval (Table 2, for all models). Based on a comparison of prior and posterior distributions, we observe that the prior distributions had minimal influence on the posterior distributions. Together, these parameters quantify the deterministic relationship between bloom size and TP load, as well as the uncertainty in this relationship, as discussed in more detail below.

The parameters of the normal and gamma models are readily comparable because both models have the same deterministic form (Figure 2). The marginal posterior distributions for β_b are well above zero for

Table 2. Posterior Parameter Means and 95% Credible Intervals for the Three Candidate Models (Parameter Units and Descriptions Same as Table 1)

	Normal Model		Gamma Model		Lognormal Model	
Symbol	Mean	95% CI	Mean	95% CI	Mean	95% CI
β_0	−19.6	−37.3–−5.21	−23.6	−39.6–−10.1	1.19	0.18–1.90
β_w	81.9	52.7–112	84.3	56.7–113	3.98	1.54–7.23
β_b	5.11	0.91–8.96	6.51	4.21–9.71	n/a	n/a
β_t	2.21	0.68–4.93	2.68	0.77–5.16	0.15	0.04–0.25
β_ψ	2.82	1.65–3.93	2.61	1.61–3.69	3.59	1.49–5.63
σ_γ	2.49	1.58–4.10	2.69	1.63–4.71	0.29	0.20–0.50
σ_ϵ	3.33	0.77–6.56	2.63	0.15–5.78	0.49	0.23–0.97

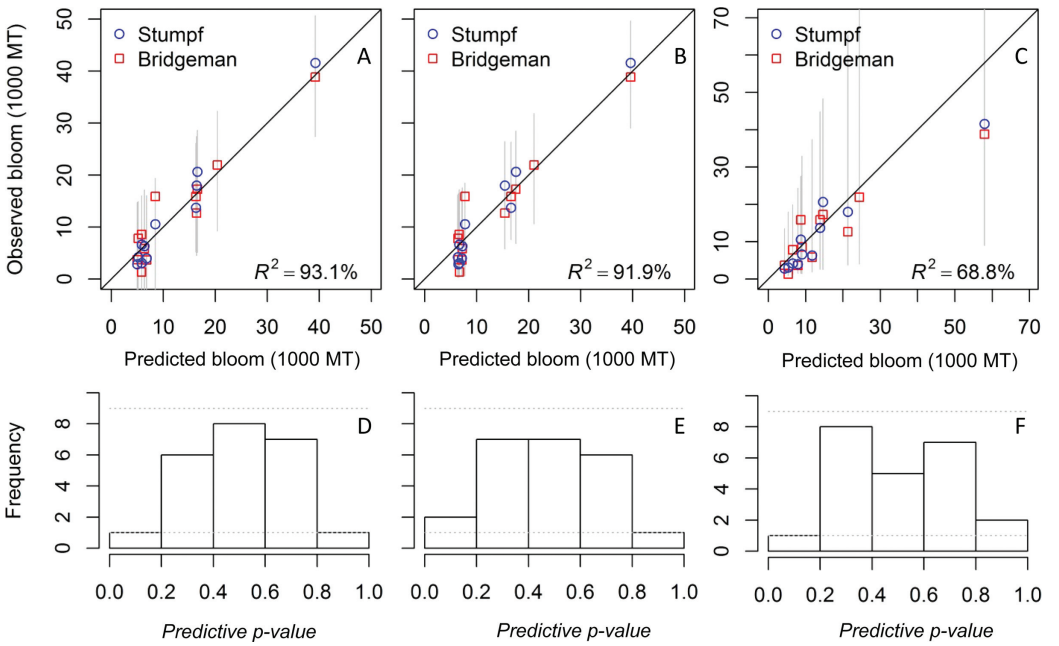

Figure 5. Full model performance: observed versus predicted bloom for the (a) normal model, (b) gamma model, and (c) lognormal model with 95% predictive intervals; and *predictive p value* distributions for the (d) normal model, (e) gamma model, and (f) lognormal model with 93% confidence bands (dashed).

both the normal and gamma models, suggesting that there is a baseline bloom level that exists even when the Maumee River nutrient loading is small. When the bloom exceeds the background level, it increases at rate β_w, which is indicated to be over 80 (MT bloom)/(MT/mo) for the normal and gamma models. The loading required for the bloom to exceed the background level is $(\beta_0+\beta_t T_i)/-\beta_w$. Thus, for 2002 ($T_i = -5$), the bloom exceeds background levels when W_i exceeds approximately 400 MT/mo, and for 2013 ($T_i = 6$), the bloom exceeds background levels when W_i exceeds approximately 100 MT/mo. (Figure 7 suggests these thresholds may be somewhat lower, reflecting the uncertainty in model parameters.) The positive estimates of β_t reflect an increase in the system's propensity for bloom formation over time. The normal and gamma model estimates of β_ψ (2.8 and 2.6, respectively) generally indicate that March–June TP load is the primary driver of cyanobacteria blooms, though February TP load contributes to a lesser degree.

The deterministic parameters for the lognormal model are more difficult to interpret due to the log transformation of the response. There is no baseline bloom parameter for the lognormal model, and the median bloom size corresponding to zero TP load is determined as $e^{(\beta_0+\beta_t T_i)}$ (Figure 2). Thus, the bloom size corresponding to zero TP load is approximately 1600 and 8100 MT for 2002 and 2013, respectively, consistent with increased sensitivity to load. The parameter β_w can be used to determine the amount of TP load required for the median bloom size to double (i.e., $\ln(2)/\beta_w$), which is approximately 170 MT/mo. The lognormal model estimate for β_ψ has a large standard deviation, indicating this parameter could not be well resolved by the model and available data.

Stochastic parameters, σ_ϵ and σ_γ, are estimated to be of similar magnitude (2500–3300 MT) for the normal and gamma models. For the lognormal model, the best estimate for $\sigma_{\gamma;L}$ was notably higher than $\sigma_{\epsilon;L}$, though their credible intervals largely overlap (Table 2). The posterior correlation between σ_ϵ and σ_γ was low for each type of model ($r^2 < 0.15$), indicating that the models provide a basis for differentiating measurement error from model prediction error.

3.2. Model Skill and Uncertainty Characterization

Model skill is assessed by examining plots of observed versus predicted values, and by determining the percent of the observation variance explained by the predictions (i.e., R^2), as shown in Figure 5. In these plots, the 95% predictive intervals account for measurement uncertainty, prediction uncertainty, and model

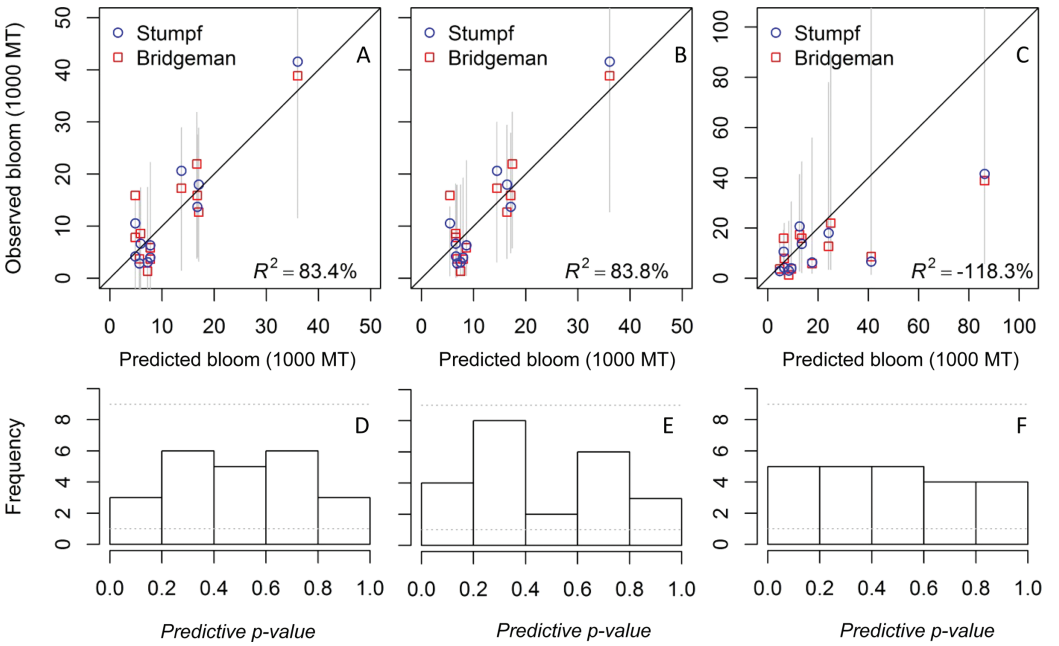

Figure 6. Model CV performance: observed versus predicted bloom for the (a) normal model, (b) gamma model, and (c) lognormal model with 95% predictive intervals; and predictive p value distributions for the (d) normal model, (e) gamma model, and (f) lognormal model with 93% confidence bands (dashed).

parameter uncertainty, such that they represent the likely range of bloom observations corresponding to each model prediction. As shown, both the normal and gamma models perform similarly, with R^2 values of 93.1% and 91.9%, respectively. The lognormal model performs less well with an R^2 value of 68.8% on the original scale. For the normal, gamma, and lognormal models, the RMSE was 2800, 3000, and 6500 MT, respectively; and the MAE was 2200, 2500, and 4500 MT, respectively.

Model uncertainty characterization is assessed by plotting histograms of predictive p values (Figures 5d–5f,). For all three models, histogram frequencies (bars) fall within the 93% confidence bands, suggesting that uncertainty is reasonably characterized. There does, however, appear to be some clustering of *p values* toward the centers of the histograms, indicating that the deviation between observations and predictions is less than expected. This discrepancy is not surprising, given that the model is being used to predict observations included in the calibration data set. We also note that the 95% predictive intervals for the normal model (Figure 5a) extend into negative space (beyond −4000 MT for some of the smaller bloom years).

3.3. Model Cross Validation

The three candidate models were also subject to cross validation (CV) to assess how well they perform when predicting observations not included within the calibration data set. The R^2 values for the normal, gamma, and lognormal models in CV mode are 83.4%, 83.8%, and −118%, respectively (Figure 6). As expected, CV performance is lower than "full model" performance (i.e., the model calibrated to all observations, as described in section 3.2) for each of the three candidate models. This decrease in performance is relatively modest for the normal and gamma models, and quite severe for the lognormal model. The predictive p value frequencies for the three models appear approximately uniformly distributed and within the 93% confidence bands, suggesting uncertainty has been well characterized (Figures 6d–6f,). While the lognormal model has substantially lower skill than the other models, the predictive distributions for the lognormal model are congruently wider (especially for large bloom predictions, due to the multiplicative error term), allowing for a reasonably uniform predictive p value distribution. For the normal, gamma, and lognormal models, RMSE was 4300, 4200, and 18,000 MT, respectively; and MAE was 3600, 3700, and 11,000 MT, respectively.

Overall, we found the gamma model, with its relatively high predictive skill and realistic uncertainty quantification, to be the preferred model. Its skill (R^2) is similar to that of the normal model and higher than that

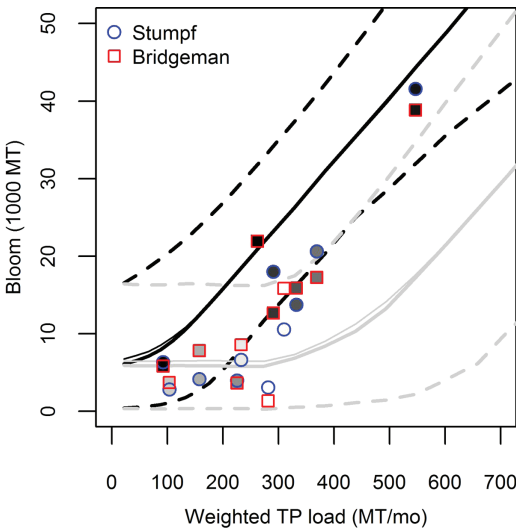

Figure 7. Relationship between observed bloom size and TP load for the gamma model in 2002 (gray lines) and 2013 (black lines), with median prediction (thick line), mean prediction (thin line), and 95% predictive intervals (dashed). Bloom observations are shaded on a linear gradient from white (2002) to black (2013).

of the lognormal model. Furthermore, we found that the difference in skill between the full model and CV results was smallest for the gamma model, suggesting it is most robust. All three models perform well in terms of uncertainty characterization based on the predictive p value distributions. However, the normal model includes predictive intervals with lower bounds <0; an undesirable result given that bloom size is a non-negative quantity.

3.4. Response Curves

The preferred (gamma) model is used to develop forecasting curves that relate the magnitude of nutrient loading to the expected range of cyanobacteria bloom observations. However, because of the temporal trend (β_t), the response curves change over time, and we illustrate the curves for 2002 and 2013 (Figure 7), representing the beginning and ending of the study period, respectively. Results suggest that a larger TP load is required to raise cyanobacteria levels beyond background levels during the early portion of the study period, while lower load is required at the end of the study period. Response curves can also be developed to show the probability of the bloom exceeding specified sizes (Figure 8).

4. Discussion

We have demonstrated an approach for quantifying uncertainty in regression-based models for harmful algal blooms. Using a flexible Bayesian framework, we were able to develop probabilistic solutions for various deterministic model forms and error distributions, allowing us to critically evaluate three alternative model formulations (using normal, lognormal, and gamma error distributions). The lognormal model, which is often an effective choice for modeling water quality data [e.g., *Ott*, 1990; *Gronewold and Borsuk*, 2010], did not perform well in this study. This is not surprising, as bloom observations are essentially aggregations of remotely sensed pixels [*Stumpf et al.*, 2012] or in situ measurements [*Bridgeman et al.*, 2013], and in accordance with the Central Limit Theorem, sufficiently large aggregations of lognormal samples are approximately normally distributed. The normal model performed well in terms of skill ($R^2 = 83.4\%$ in CV), but its predictive distributions did include unrealistic negative values, especially when the predicted bloom size was small. The gamma model (i.e., the model that uses gamma distributions to represent stochastic error terms, γ and ϵ) avoids negative predictive values and its predictive distributions approach normality as predicted bloom size increases [e.g., *Thom*, 1958]. While gamma models have been used to characterize water quality sampling data

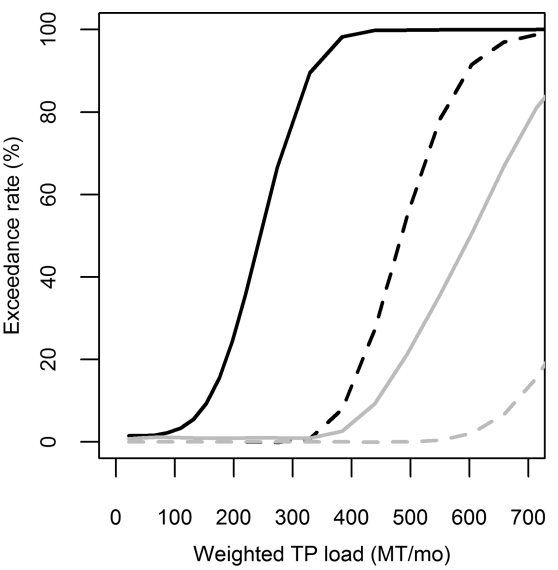

Figure 8. Probability of exceeding 20,000 MT bloom (solid) and 40,000 MT bloom (dashed) for a range of loads under 2002 conditions (gray) and 2013 conditions (black).

[e.g., *Gilliom and Helsel*, 1986], this is the first study that we are aware of to use the gamma distribution for characterizing predictive uncertainty in a water quality forecasting model. Because the gamma distribution realistically characterized uncertainty, while also maintaining high predictive skill ($R^2 = 83.8\%$ in CV), it was found to be the most preferred model. In addition, the gamma (and normal) model is advantageous relative to the lognormal model because it is primarily based on linear relationships that are interpreted more easily than the nonlinear relationships imposed by the lognormal model formulation [*Faraway*, 2005]. The lognormal model formulation necessitates that bloom size increases exponentially with TP load (Figure 2), which is implausible outside of a limited range of TP loads, due to stoichiometric limitations on the amount of cyanobacteria that can be produced per unit load.

The Bayesian modeling approach also allowed us to represent multiple sources of stochasticity through a hierarchical formulation. For comparison, the gamma model was also developed omitting the yearly random effect term, γ, such that all error was represented by ϵ. In this case, all observations are treated as independent (i.e., intraclass correlation is ignored), which is incorrect given that there are two related observations for each year (see section 2.3). The predictive performance of this model ($R^2 = 92.3\%$, RMSE = 2900, and MAE = 2400) is similar to that of the original gamma model. However, by omitting γ, the credible intervals for deterministic model parameters are reduced by 24%, on average, suggesting we have more confidence in how the system functions than is really the case. This, in turn, results in a forecasting curve with 13% narrower predictive intervals, on average.

Clearly, rather broad uncertainty bounds need to be incorporated in the simple empirical models currently employed to relate large-scale hydrological and pollutant loading variables to bloom size [e.g., *Stumpf et al*. 2012; *Ohio EPA*, 2013; *IJC*, 2014]. Without uncertainty quantification, forecast users may not adequately prepare for worst-case scenarios (i.e., when blooms are substantially larger than the forecasting model's best estimate), and prediction-observation discrepancies may lead to model skepticism and mismanagement of the system. In the future, it may be reasonable to include additional predictor variables within the model to reduce predictive uncertainty. This may be justified by a larger calibration data set, as additional bloom observations become available, and by process-based studies that help quantify important factors controlling bloom size. Additional predictors can be readily incorporated into the probabilistic models developed here, and their forecasting efficacy can be readily evaluated using the procedures outlined in this study.

Our model is based on certain assumptions which could potentially be modified in the future, as additional and more refined bloom observations become available. In this study, we assume that both sets of observations provide equally uncertain representations of the true bloom size. As such, they are assigned the same measurement error variance, σ_ϵ^2, which is determined through model calibration. Assuming equal measurement error variance is reasonable, given that both sets of observations are based on peer-reviewed studies, and both have known limitations. For example, satellite remote sensing only measures cyanobacteria near the water surface [*Wynne et al*., 2010; *Stumpf et al*., 2012]; and while phytoplankton tows sample throughout the entire water column, they have relatively limited spatial and temporal resolution [*Bridgeman et al*., 2013]. If, in the future, the measurement errors associated with these two data sets could be characterized independently, then it would be possible to incorporate these uncertainties directly into the model, effectively giving more weight to observations with lower measurement uncertainty. It should also be noted that the two bloom data sets provide relative, rather than absolute, measures of bloom size. As described in section 2.1, neither of the data sets could be converted to mass units without making multiple assumptions. Thus, our modeling predictions are also largely relative in nature, and the additional uncertainty associated with the conversion to mass units is not addressed explicitly in this study. If, in the future, independent observations of bloom mass become available, then this information could be used to more rigorously characterize the absolute bloom mass and its uncertainty.

This study implies a linkage between TP load and bloom size. For comparison, forecasting models were also developed using DRP and nitrate (in place of TP) which explain 88.0% and 72.8% of the variability in bloom size, respectively (compared to $R^2 = 91.9\%$ for TP). These results suggest that TP is the most effective bloom predictor, consistent with our expectation. Although nitrogen limitation may occur in some circumstances [*Chaffin et al*., 2013], phosphorus is generally acknowledged to be the primary limiting nutrient in Lake Erie [*Guildford et al*., 2005; *Wilhelm et al*., 2003; *Steffen et al*., 2014]. In addition, while DRP is often considered highly bioavailable [e.g., *Michalak et al*., 2013], many non-DRP species of phosphorus are also known to be bioavailable [*Li and Brett*, 2013]. Furthermore, the speciation of phosphorus measured in the Maumee River

in spring may not be representative of phosphorus speciation in the lake during the late-summer bloom. As such, TP load, rather than DRP load, may be the best surrogate for bioavailable phosphorus at the time of the late-summer bloom.

Our results suggest that Lake Erie's western basin has become increasingly susceptible to large cyanobacteria blooms over the last 12 years, independent of the spring nutrient load. For all three candidate TP models, the temporal trend parameter, β_t, was found to be significantly positive (i.e., the lower bound of each 95% credible interval was greater than zero, Table 2). The trend parameters were also significantly positive for the alternative models using DRP and nitrate in place of TP. Based on the preferred TP model, the threshold loading rate for the bloom to exceed background levels has dropped substantially between 2002 and 2013 (Figure 7). While the cause of this change cannot be determined from this empirical study, we note that the increase in susceptibility is consistent with *Michalak et al*. [2013], who suggested that large blooms may be increasingly common in the future due to changing meteorological conditions that promote cyanobacteria growth. However, late-summer Lake Erie water surface temperature has not increased significantly over the study period (NOAA, http://coastwatch.glerl.noaa.gov/glsea/doc/), suggesting that warming is not the primary cause of larger blooms. Future research could examine whether other climate change factors, such as increasingly calm summer conditions [*Michalak et al*., 2012], may better explain this increasing trend in bloom susceptibility. An increasing reservoir of *Microcystis* seed colonies within Lake Erie's benthos is another possible cause of increasing cyanobacteria blooms [*Rinta-Kanto et al*., 2009]. It may also be that invasive species, such as dreissenid mussels, have promoted *Microcystis* dominance over time through selectively feeding on competing phytoplankton species [*Vanderploeg et al*., 2001] and through alteration of the phosphorus cycle [*Steffen et al*., 2014]. USGS *Dreissena* surveys for western Lake Erie suggest a decrease in zebra mussel abundance, but an increase in quagga and total dreissenid mussel abundance over the last decade [*EC and U.S. EPA*, 2014].

The relationships between bloom size and Maumee River TP load (Figures 7 and 8) can be used to inform cyanobacteria bloom reduction plans. In the model, the weighted average TP load is a function of β_ψ, as described in section 2.4. For the gamma model, β_ψ is approximately 2.6, which implies that March through June loads receive full weight, February loads receive partial weight, and January loads receive negligible weight. For the study period, weighted spring TP loads varied from 98 to 554 MT/mo, with a mean of 270 MT/mo. This is largely consistent with *Stumpf et al*. [2012], who found that March–June TP loads (averaging 272 MT/mo over the study period) were an effective predictor of bloom size. Task force reports by *Ohio EPA* [2013] and the International Joint Commission [*IJC*, 2014] both recommend spring (March through June) load targets of 200 MT/mo. According to the IJC report, this loading target would reduce the average bloom size to 1.0 unit of CI (i.e., 4800 MT, per section 2.1). However, based on this study (Figure 7), a target load of 200 MT/mo is expected to result in a mean bloom of approximately 16,200 (95% credible interval: 8500–23,000) MT under current 2013 conditions, due to the lake's apparent increasing susceptibility to blooms. Under earlier 2002 conditions, the target load would be expected to produce an average bloom size of approximately 6600 (95% credible interval: 4200–9700) MT, more similar to the IJC prediction.

It is currently unknown whether Lake Erie's susceptibility to cyanobacteria blooms will continue to increase over time, and this uncertainty complicates our ability to develop scenarios for the effects of loads in future years. However, we note that the temporal trend has been gradual, such that the year-to-year change in bloom susceptibility is not likely to be great. Thus, we suggest making following year predictions using the 2013 load-bloom relationships (Figures 7 and 8), and updating the analysis with additional bloom observations as they become available.

5. Summary

This study demonstrates the efficacy of a Bayesian hierarchical framework for empirical modeling and forecasting of cyanobacteria blooms. The flexibility of this framework allows us to represent predictive and measurement errors as gamma probability distributions, and the resulting model is found to have relatively high predictive skill and realistic uncertainty quantification. An otherwise equivalent model using a normal error distribution is found to perform similarly, but it includes unrealistic negative bloom values within its predictive distribution, especially when the predicted bloom size is small. A model using a log-transformed response (and lognormal error distribution) is shown to have relatively poor predictive skill, perhaps in part

because an exponential relationship between bloom size and TP load defies stoichiometric limitations on bloom production. In addition, the study demonstrates how hierarchical modeling can be used to explicitly represent both model and measurement error, when there are multiple (imperfect) measurements of each bloom. This, in turn, prevents underestimation of model parameter uncertainty, relative to treating all bloom measurements as statistically independent. While this study focuses on western Lake Erie cyanobacteria blooms, the approaches developed here should be readily transferable to other freshwater and coastal systems where pollutant loads produce measurable large-scale impacts.

From a management perspective, this study supports the proposition that spring TP load is an effective (i.e., significant) predictor of Lake Erie cyanobacteria bloom size [*Stumpf et al*., 2012; *Ohio EPA*, 2013; *IJC*, 2014]. However, there is substantial uncertainty in the TP load-bloom relationship, and this uncertainty should be communicated when preparing bloom forecasts and setting management goals. Results also suggest that Lake Erie has become increasingly susceptible to large cyanobacteria blooms over the study period (2002–2013), potentially complicating management efforts to reduce the bloom size to target levels.

Acknowledgments
This project was funded by the Great Lakes Restoration Initiative (administered by the U.S. EPA) Synthesis, Observation, and Response (SOAR) project, the University of Michigan Water Center, the Cooperative Institute for Limnology and Ecosystems Research (CILER), and the National Oceanic and Atmospheric Administration (NOAA). We thank Richard Stumpf and Thomas Bridgeman for providing bloom data and manuscript comments, and Steven Ruberg for providing manuscript review. In addition, we thank Peter Richards and David Baker for assistance with the loading data. All data sources are described in sections 2.1 and 2.2. This is NOAA-GLERL contribution 1719.

References

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*(8), 1393–1442.
Bertram, P. E. (1993), Total phosphorus and dissolved oxygen trends in the central basin of Lake Erie, 1970–1991, *J. Great Lakes Res.*, *19*, 224–236.
Bridgeman, T. B., J. D. Chaffin, D. D. Kane, J. D. Conroy, S. E. Panek, and P. M. Armenio (2012), From river to lake: Phosphorus partitioning and algal community compositional changes in Western Lake Erie, *J. Great Lakes Res.*, *38*, 90–97.
Bridgeman, T. B., J. D. Chaffin, and J. E. Filbrun (2013), A novel method for tracking western Lake Erie *Microcystis* blooms, 2002–2011, *J. Great Lakes Res.*, *39*(1), 83–89.
Brooks, S. P., and A. Gelman (1998), General methods for monitoring convergence of iterative simulations, *J. Comput. Graph. Stat.*, *7*(4), 434–455.
Burns, N. M., D. C. Rockwell, P. E. Bertram, D. M. Dolan, and J. J. Ciborowski (2005), Trends in temperature, Secchi depth, and dissolved oxygen depletion rates in the central basin of Lake Erie, 1983–2002, *J. Great Lakes Res.*, *31*, 35–49.
Cha, Y., S. S. Park, K. Kim, M. Byeon, and C. A. Stow (2014), Probabilistic prediction of cyanobacteria abundance in a Korean reservoir using a Bayesian Poisson model, *Water Resour. Res.*, *50*, 2518–2532, doi:10.1002/2013WR014372.
Chaffin, J. D., T. B. Bridgeman, and D. L. Bade (2013), Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie, *Adv. Microbiol.*, *3*(6), 16–26.
Charlton, M. N., J. E. Milne, W. G. Booth, and F. Chiocchio (1993), Lake Erie offshore in 1990—Restoration and resilience in the central basin, *J. Great Lakes Res.*, *19*(2), 291–309.
Chatfield, C. (2006), *Model Uncertainty. Encyclopedia of Environmetrics*, John Wiley, Hoboken, N. J.
Clark, J. S. (2007), *Models for Ecological Data: An Introduction*, Princeton Univ. Press, Princeton, N. J.
Conroy, J. D., D. D. Kane, and D. A. Culver (2008), Declining Lake Erie ecosystem health-evidence from a multi-year, lake-wide, plankton study, in *Checking the Pulse of Lake Erie*, edited by M. Munawar and R. Heath, pp. 369–408, Aquat. Ecosyst. Health and Manage. Soc., Burlington, Ont.
Cooley, D., and S. R. Sain (2010), Spatial hierarchical modeling of precipitation extremes from a regional climate model, *J. Agric. Biol. Environ. Stat.*, *15*(3), 381–402.
Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle (2009), Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling, *Ecol. Appl.*, *19*(3), 553–570.
Davis, C. C. (1964), Evidence for the eutrophication of Lake Erie from phytoplankton records, *Limnol. Oceanogr.*, *9*(3), 275–283.
DePinto, J. V., T. C. Young, and L. M. McIlroy (1986), Impact of phosphorus control measures on water quality of the Great Lakes, *Environ. Sci. Technol.*, *20*, 752–759.
Dolan, D. M. (1993), Point source loadings of phosphorus to Lake Erie: 1986–1990, *J. Great Lakes Res.*, *19*, 212–223.
Environment Canada (EC) and U.S. EPA (2014), *State of the Great Lakes 2011*, Environ. Can. and the U.S. Environ. Prot. Agency, Ottawa, Ont.
Elmore, K. L. (2005), Alternatives to the chi-square test for evaluating rank histograms from ensemble forecasts, *Weather Forecasting*, *20*(5), 789–795.
Elsner, J. B., and C. P. Schmertmann (1994), Assessing forecast skill through cross validation, *Weather Forecasting*, *9*(4), 619–624.
Faraway, J. J. (2005), *Linear Models with R*, Chapman and Hall, Boca Raton, Fla.
Gelman, A. (2006), Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Anal.*, *1*(3), 515–534.
Gelman, A., and J. Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge Univ. Press, Cambridge, U. K.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, Chapman and Hall, Boca Raton, Fla.
Gilliom, R. J., and D. R. Helsel (1986), Estimation of distributional parameters for censored trace level water quality data: 1. Estimation techniques, *Water Resour. Res.*, *22*(2), 135–146.
Gronewold, A. D., and M. E. Borsuk (2010), Improving water quality assessments through a hierarchical Bayesian analysis of variability, *Environ. Sci. Technol.*, *44*(20), 7858–7864.
Gronewold, A. D., and C. A. Stow (2014), Unprecedented seasonal water level dynamics on one of the earth's largest lakes, *Bull. Am. Meteorol. Soc.*, *95*(1), 15–17.
Gronewold, A. D., S. S. Qian, R. L. Wolpert, and K. H. Reckhow (2009), Calibrating and validating bacterial water quality models: A Bayesian approach, *Water Res.*, *43*(10), 2688–2698.
Guildford, S. J., R. E. Hecky, R. E. Smith, W. D. Taylor, M. N. Charlton, L. Barlow-Busch, and R. L. North (2005), Phytoplankton nutrient status in Lake Erie in 1997, *J. Great Lakes Res.*, *31*, 72–88.
Härdle, W., A. Werwatz, M. Müller, and S. Sperlich (2004), *Nonparametric and Semiparametric Models*, Springer, Berlin.

Hawley, N., T. H. Johengen, Y. R. Rao, S. A. Ruberg, D. Beletsky, S. A. Ludsin, B. J. Eadie, D. J. Schwab, T. E. Croley, and S. B. Brandt (2006), Lake Erie hypoxia prompts Canada-US study, *Eos Trans. AGU*, *87*, 313–319.
International Joint Community (1978), *Great Lakes Water Quality Agreement of 1978, Agreement with Annexes and Terms of Reference*, Between the United States of America and Canada, Windsor, Ont.
International Joint Community (2014), A balanced diet for Lake Erie: Reducing phosphorus loadings and harmful algal blooms, report of the lake Erie ecosystem priority, Windsor, Ont. [Available at http://www.ijc.org/en_/leep/report.]
Li, B., and M. T. Brett (2013), The influence of dissolved phosphorus molecular form on recalcitrance and bioavailability, *Environ. Pollut.*, *182*, 37–44.
Ludsin, S. A., M. W. Kershner, K. A. Blocksom, R. L. Knight, and R. A. Stein (2001), Life after death in Lake Erie: Nutrient controls drive fish species richness, rehabilitation, *Ecol. Appl.*, *11*, 731–746.
Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000), WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility, *Stat. Comput.*, *10*, 325–337.
Makarewicz, J. C. (1993), Phytoplankton biomass and species composition in Lake Erie, 1970 to 1987, *J. Great Lakes Res.*, *19*, 258–274.
Makarewicz, J. C., and P. Bertram (1991), Evidence for the restoration of the Lake Erie ecosystem, *Bioscience*, *41*(4), 216–223.
Makarewicz, J. C., T. Lewis, and P. Bertram (1989), *Phytoplankton and Zooplankton Composition, Abundance and Distribution and Trophic Interactions: Offshore Region of Lake Erie, Lake Huron, and Lake Michigan*, 1985, U.S. EPA, Chicago, Ill.
Malve, O., M. Laine, H. Haario, T. Kirkkala, and J. Sarvala (2007), Bayesian modelling of algal mass occurrences—Using adaptive MCMC methods with a lake water quality model, *Environ. Modell. Software*, *22*(7), 966–977.
Meng, X. L. (1994), Posterior predictive p-values, *Ann. Stat.*, *22*(3), 1142–1160.
Michalak, A. M., et al. (2013), Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions, *Proc. Natl. Acad. Sci. U. S. A.*, *110*(16), 6448–6452.
Mortimer, C. H. (1987), Fifty years of physical investigations and related limnological studies on Lake Erie, 1928–1977, *J. Great Lakes Res.*, *13*(4), 407–435.
Najafi, M. R., and H. Moradkhani (2013), Analysis of runoff extremes using spatial hierarchical Bayesian modeling, *Water Resour. Res.*, *49*, 6656–6670, doi:10.1002/wrcr.20381.
Nash, J., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.
Obenour, D. R., A. M. Michalak, and D. Scavia (2014), Assessing biophysical controls on Gulf of Mexico hypoxia through probabilistic modeling, *Ecol. Appl.*, in press, doi:10.1890/13-2257.1.
Ohio EPA. (2013), Ohio Lake Erie phosphorus task force II, final report, Ohio Environmental Protection Agency, Columbus, Ohio. [Available at http://www.epa.state.oh.us/portals/35/lakeerie/ptaskforce2/Task_Force_Report_October_2013.pdf.]
Ott, W. R. (1990), A physical explanation of the lognormality of pollutant concentrations, *J. Air Waste Manage. Assoc.*, *40*(10), 1378–1383.
Qian, S. S., T. F. Cuffney, I. Alameddine, G. McMahon, and K. W. Reckhow (2010), On the application of multilevel modeling in environmental and ecological studies, *Ecology*, *91*(2), 355–361.
R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, Vienna. [Available at http://www.R-project.org.]
Rinta-Kanto, J. M., E. A. Konopko, J. M. DeBruyn, R. A. Bourbonniere, G. L. Boyer, and S. W. Wilhelm (2009), Lake Erie *Microcystis*: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake, *Harmful Algae*, *8*(5), 665–673.
Robson, B. J. (2014), When do aquatic systems models provide useful predictions, what is changing, and what is next?, *Environ. Modell. Software*, in press, doi:10.1016/j.envsoft.2014.01.009.
Rucinski, D., D. Scavia, J. DePinto, and D. Beletsky (2014), Modeling Lake Erie's hypoxia response to nutrient loads and physical variability, *J. Great Lakes Res.*, in press, doi:10.1016/j.jglr.2014.02.003.
Scavia, D., et al. (2014), Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia, *J. Great Lakes Res.*, *40*(2), 226–246.
Steffen, M. M., B. S. Belisle, S. B. Watson, G. L. Boyer, and S. W. Wilhelm (2014), Status, causes and controls of cyanobacterial blooms in Lake Erie, *J. Great Lakes Res.*, *40*(2), 215–225.
Stumpf, R. P., T. T. Wynne, D. B. Baker, and G. L. Fahnenstiel (2012), Interannual variability of cyanobacterial blooms in Lake Erie, *PLoS One*, *7*, e42444.
Sturtz, S., U. Ligges, and A. E. Gelman (2005), R2WinBUGS: A package for running WinBUGS from R, *J. Stat. Software*, *12*(3), 1–16.
Thom, H. C. (1958), A note on the gamma distribution, *Mon. Weather Rev.*, *86*(4), 117–122.
Vanderploeg, H. A., J. R. Liebig, W. W. Carmichael, M. A. Agy, T. H. Johengen, G. L. Fahnenstiel, and T. F. Nalepa (2001), Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie, *Can. J. Fish. Aquat. Sci.*, *58*(6), 1208–1221.
Westfall, P., and K. S. Henning (2013), *Understanding Advanced Statistical Methods*, CRC Press, Boca Raton, Fla.
Wilhelm, S. W., J. M. DeBruyn, O. Gillor, M. R. Twiss, K. Livingston, R. A. Bourbonniere, L. D. Pickell, C. G. Trick, A. L. Dean, and R. M. L. McKay (2003), Effect of phosphorus amendments on present day plankton communities in pelagic Lake Erie, *Aquat. Microb. Ecol.*, *32*(3), 275–285.
Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, and J. Dyble (2010), Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data, *Limnol. Oceanogr.*, *55*(5), 2025–2036.
Zhou, Y., D. R. Obenour, D. Scavia, T. H. Johengen, and A. M. Michalak (2013), Spatial and temporal trends in Lake Erie hypoxia, 1987–2007, *Environ. Sci. Technol.*, *47*, 899–905.