Environmental Science & Technology

Article

pubs.acs.org/est

Quantifying and Reducing Uncertainty in Estimated Microcystin Concentrations from the ELISA Method

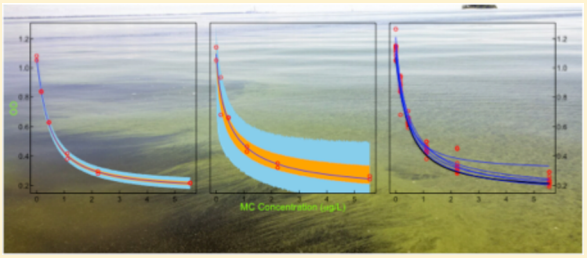
Song S. Qian,*[,†] Justin D. Chaffin,[‡] Mark R. DuFour,[†] Jessica J. Sherman,[†] Phoenix C. Golnick,[†] Christopher D. Collier,[†] Stephanie A. Nummer,[†] and Michaela G. Margida[†]

[†]Department of Environmental Sciences, The University of Toledo, Toledo, Ohio 43606, United States

[‡]Franz Theodore Stone Laboratory, The Ohio State University, Put-in-Bay, Ohio 43456, United States

Supporting Information

ABSTRACT: We discuss the uncertainty associated with a commonly used method for measuring the concentration of microcystin, a group of toxins associated with cyanobacterial blooms. Such uncertainty is rarely reported and accounted for in important drinking water management decisions. Using monitoring data from Ohio Environmental Protection Agency and from City of Toledo, we document the sources of measurement uncertainty and recommend a Bayesian hierarchical modeling approach for reducing the measurement uncertainty. Our analysis suggests that (1) much of the uncertainty is a result of the highly uncertain "standard curve" developed during each test and (2) the uncertainty can be reduced by pooling raw test data from multiple tests. Based on these results, we suggest that estimation uncertainty can be effectively reduced through the effort of either (1) regional regulatory agencies by sharing and combining raw test data from regularly scheduled microcystin monitoring program or (2) the manufacturer of the testing kit by conducting additional tests as part of an effort to improve the testing kit.



INTRODUCTION

Cyanobacterial blooms have become a global problem due to excessive anthropogenic nutrient inputs and a warming climate.[1,2] Aside from the negative ecological impacts associated with blooms,[3] these blooms also pose a threat to human health because they produce toxic compounds that impair the nervous system, liver, and skin.[4] Because toxic cyanobacterial blooms often occur in waters that are sources for drinking water, frequent and accurate quantification of cyanboacterial toxins in treatment-finished drinking water is paramount in protecting the public. Microcystins (MCs) are a group of liver toxins commonly produced by many genera of cyanobacteria worldwide.[5] These toxins have been associated with liver cancers and human fatalities; particularly in people with poor liver function already being treated through dialysis.[4,6] Although many authors have discussed the need for controlling harmful algal blooms (through reducing N and P loadings to waters) to reduce the risk of MC exposure, the problem of MC measurement uncertainty is not adequately discussed in the literature. Information about measurement uncertainty is important to local government agencies responsible for providing safe drinking water.

The Toledo water crisis of August 2014 started when one MC concentration measurement from a treated water sample in Collins Park Water Treatment Plant of City of Toledo, Ohio, exceeded the Ohio drinking water standard of 1 μg/L. (The Ohio standard is based on the World Health Organization drinking water quality criterion for microcystin.[7]) After the initial detection, three additional tests, each with the usual two replicates, were carried out on the same day using the same water sample and each time at least one replicate showed a concentration above the criterion. These results prompted the City of Toledo to issue a "Do not drink" advisory on the morning of August 2, 2014, affecting about a half million residents. Additional tests were conducted on drinking water from the water treatment plant and throughout the distribution system until all samples consistently showed microcystin concentration below "detectable" levels (<0.30 μg/L) during the water advisory, which lasted nearly 3 days. In the aftermath of the crisis, some questioned the wisdom of issuing the advisory based on one sample exceeding the criterion.[8] An important question raised by this incident is how to properly communicate the risk of drinking water contamination to the public?

The City of Toledo's source water is from western Lake Erie, which has been plagued by dense blooms of the microcystin-producing cyanobacterium *Microcystis* since the early 2000s.[9−11] Projected future meteorological and agricultural conditions are expected to promote large cyanobacterial blooms in Lake Erie.[12]

While the harmful effects of microcystin have been reported widely,[13] the risk of exposure to harmful levels of the toxins has not been adequately communicated. This is largely because of the lack of information on the uncertainty associated with concentrations measured using common laboratory equipment.

Received: June 23, 2015
Revised: October 27, 2015
Accepted: October 30, 2015

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

In this paper, we describe a statistical modeling approach for quantifying such uncertainty. We analyzed data from 21 tests conducted at the Franz Theodore Stone Laboratory of the Ohio State University (Stone Lab) during the 2014 cyanobacterial bloom season and proposed a Bayesian hierarchical modeling approach for reducing the estimation uncertainty.

METHODOLOGICAL BACKGROUND

The ELISA Method. The enzyme-linked immunosorbent assay (ELISA) method[14,15] is used for measuring microcystin (MC) concentration in almost all Ohio drinking water facilities that use Lake Erie as source water. The ELISA method quantifies microcystin concentrations through a competitive binding process between toxins and enzyme-labeled microcystin in the inside wall of test wells. The toxin concentration is visualized with a color development process and is inversely proportional to color development (i.e., the darker the sample color, the less toxin present) with a nonlinear relationship. During a test, a number of samples with known toxin concentrations (standard solutions) were used to develop a "standard curve", a mathematical model of optical density (OD) as a function of toxin concentrations. The ELISA test kit used by Stone Lab and Toledo Water Department (Abraxis #520011, Warminster, PA) has six standard concentrations (0, 0.15, 0.40, 1.00, 2.00, and 5.00 μg/L used in Stone Lab and 0, 0.167, 0.444, 1.110, 2.220, and 5.550 used in Toledo Water Department). Two replicates were used in both facilities, resulting in 12 measured ODs for developing a standard curve for each test. Water samples with unknown MC concentrations are put in the remaining wells of the 96-well plate in the same test and their MC concentrations are estimated (predicted) by the fitted standard curve.

Statistical Concerns. The process of measuring MC concentration has two steps: (1) developing a standard curve using solutions with known MC concentrations and (2) estimating the unknown MC concentrations in water samples. Statistical issues arise in three areas: model specification, sample size, and predictive uncertainty.

Because we do not know the underlying "true" relationship between OD and MC concentration, we consider any specific model form as an approximation. Although model choice includes considering alternative models, we limit our discussion to the choice of response and predictor variables. Because a standard curve is inevitably quantified using regression, the choice of the regression model response variable determines the accuracy of the measured MC concentration. When conducting a test, we know the MC concentrations of standard solutions and measure the respective ODs. Conceptually, the response variable should be OD and the predictor variable should be MC concentration. The resulting regression model, however, is fit to minimize the predictive error of OD, not MC concentration. When using the resulting model for estimating MC concentrations, the estimation uncertainty will be larger than the model summary statistics would suggest.

The City of Toledo uses a 4-parameter nonlinear regression model, recommended by the kit manufacturer (Abraxis), with OD as the response variable:

$$\mathrm{OD} = \frac{\alpha_1 - \alpha_4}{1 + \left(\frac{c_{st}}{\alpha_3}\right)^{\alpha_2}} + \alpha_4 + \varepsilon \tag{1}$$

where ε is the model error term (and $\varepsilon \sim N(0, \sigma^2)$), $\alpha_1 - \alpha_4$ are unknown regression coefficients to be estimated, OD is the observed OD value, and c_{st} is the standard solution concentration. As in most empirical models, regression coefficients do not have physical meanings. When model coefficients are estimated, the fitted model is optimized to reduce prediction error in OD. Using the inverse model of eq 1 for estimating MC concentration will lead to larger than expected estimation uncertainty (based on regression model summary statistics such as residual variance and R^2) because model coefficients are estimated to minimize the prediction error in OD, not in c_{st} (see Results for details).

The Stone Lab uses another manufacturer-recommended standard curve model, which is a log−linear regression model after transforming the measured OD values to relative ODs—the measured ODs from standard solutions with nonzero MC concentrations and water samples with unknown MC concentrations were divided by the average of the two OD values from the two standard solutions with a concentration of 0.0 (the "%B/B$_0$" term specified by Abraxis #520011 instructions):

$$\log(c_{st}) = \beta_0 + \beta_1 r\mathrm{OD} + \varepsilon \tag{2}$$

where rOD is the relative OD. The resulting standard curve is optimized for predicting log MC concentration. The transformation of OD to rOD requires averaging measured OD values from two replicates, which reduces the uncertainty in the predictor variable, but also reduces sample size of the regression model.

The second area of concern is related to the sample size used for quantifying the standard curve. The standard curve is typically developed based on a small number of standard solutions, that is, the regression model is developed based on a small sample size. As a result, the estimated standard curve is inherently variable. However, the quality of a fitted standard curve is commonly judged only by one or two summary statistics, most likely the coefficient of determination or R^2 value and residual sum of squares. (The R^2 of the nonlinear regression model in eq 1 is calculated in the software of Abraxis kit to be 1 minus the ratio of residual variance over the sample variance of response variable data.) We note that the R^2 statistic is not commonly associated with nonlinear regression models. When fit to data from a typical ELISA test kit (six standard concentrations), the model of eq 1 has a degrees of freedom of 8 (six observations with two replicates and four model parameters) and the model of eq 2 has a degrees of freedom of 3 (five observations associated with nonzero concentrations and two replicates were averaged before curve fitting). We note that replicate standard solutions are not independent samples, as a result, the degrees of freedom of the model of eq 1 is smaller than 8 (and the uncertainty estimated based on a degrees of freedom of 8 represents an underestimate).

A regression model with a small degrees of freedom is often characterized by a near-perfect R^2 value. A small degrees of freedom implies that each individual data point can have relatively large influence on the fitted standard curve. As long as measurement error (in OD) is present, we expect a relatively high variation in estimated coefficients from test to test. For a given test, the resulting standard curve may fit to the data very well, but we cannot put much confidence in it because a very different curve is likely when we conduct the same test the next time.

An extreme example is to fit a linear model using two data points. The fitted line is always perfect without error (with $R^2 = 1$). This is because the fitted model includes the observation error in the two data points (hence a perfect R^2 value).

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

As a result, our confidence in the fitted line being the underlying true model is always low. (If we repeat the model fitting process with two new data points from the same population, we expect to see a different model with perfect model diagnostic statistics.) Furthermore, we cannot evaluate the fitted model using data alone because we exhausted information in the two data points to estimate the two parameters of the linear model. In this regard, degrees of freedom of a regression model is a measure of information available for model assessment using statistics. A small degrees of freedom, therefore, indicates a lack of confidence in the model no matter how well the model fits to the data. In general, the smaller the sample size, the more likely a fitted regression model will represent idiosyncrasies specific to the data used for fitting the model (noise). Therefore, the model is more likely to have a small residual variance and a high R^2 value, but less likely to be able to predict future cases. This is why we can often improve a regression model's summary statistics by fitting the model using aggregated data (e.g., means of replicates), but without improving the model's predictive capability.

The third area of concern is the predictive uncertainty. Once the standard curve is developed, we estimate MC concentrations of water samples based on their OD values. The predictive uncertainty associated with the established standard curve is often ignored by treating the estimated standard curve as known and present the expected concentration value without uncertainty information.

Finally, we note that when the estimated concentration value is below the lowest nonzero standard solution concentration (e.g., 0.15 μg/L) or larger than the largest concentration (e.g., 5 μg/L), it is considered as censored and reported as <0.15 or >5 μg/L, respectively. But the estimation process does not have a "method reporting limit." Rather, this practice is to avoid extrapolating a regression model for prediction, due to the uncertainty in the model form, as well as the increased predictive variance when predicting points beyond the range of data used to fit the model.

Regression Model Predictive Uncertainty. A regression model's predictive uncertainty can be presented as the predictive distribution of a future observation. In the context of the ELISA method, this predictive distribution is the probability distribution of the unknown MC concentration of a water sample given the observed OD value (π(MC|OD), where π represent a density function). For example, when using the log−linear model in (2), the log transformed MC concentration is assumed to be a normal distribution with mean quantified by the linear regression model and the variance estimated by the residual variance. Consequently, this predictive distribution is expressed in terms of regression model coefficients, that is, $\pi(\log(\mathrm{MC})\,|r\mathrm{OD}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$, which is normal with mean $\hat{\beta}_0 + \hat{\beta}_1 r\mathrm{OD}$ and variance $\hat{\sigma}^2$.

The estimated model coefficients ($\hat{\beta}_0$, $\hat{\beta}_1$) and sample variance of model residuals ($\hat{\sigma}^2$) are themselves random variables. As a result, uncertainty in these estimated quantities must be considered. To simplify the notation, we use θ to represent the vector of model coefficients. A general expression of the predictive distribution is

$$\pi(\log(\mathrm{MC})|r\mathrm{OD}) = \int \pi(\log(\mathrm{MC})|r\mathrm{OD}, \hat{\theta}, \hat{\sigma}^2) \times \pi(\hat{\theta}|\hat{\sigma}^2) \times \pi(\hat{\sigma}^2) d\hat{\sigma}^2 d\hat{\theta} \quad (3)$$

The analytic result of (3) is available for a linear model.[16] For a nonlinear regression model such as eq 1, a general analytic solution is not available. However, numerical integration using Monte Carlo simulation can be used to approximate the predictive distribution.

Because the primary model in our study is a log−linear regression model, retransformation bias[17] is a practical concern. A Monte Carlo simulation based approach has many advantages over alternative methods,[18] especially when we are interested in more than just a bias correction for the predicted means. Qian[19] (Chapter 9) described a general purpose Monte Carlo simulation algorithm to approximate the integral of eq 3, both for linear and nonlinear regression models (see online Supporting Information).

Bayesian Hierarchical Modeling. Reviewing standard curves developed in Stone Lab in the past, we find that fitted standard curves not only have a considerable within-test variation, but also vary from test to test. On the one hand, we may be tempted to interpret the among-test variation as a result of inherent systematic (albeit unknown) differences among tests, thereby, justifying the practice of estimating a different standard curve for every test. On the other hand, a large *random* among-test variation is expected because of the small degrees of freedom, which requires us to pool data from multiple tests to improve the model.

These two opposing approaches for modeling the data can be unified by combining data from multiple tests using a Bayesian hierarchical model (BHM), in which data from different tests are treated as different, but their distribution models are allowed to share common features. Using the model in eq 2 as an example, the hierarchical modeling approach sets the model at the test level using the same log−linear model as in eq 2, but connects the models for different tests by imposing a common prior distribution of model coefficients:

$$\begin{aligned} \log(c_{st,\mathrm{ij}}) &= \beta_{0j} + \beta_{1j} r\mathrm{OD}_{ij} + \varepsilon_{ij} \\ \begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} &\sim \mathrm{MVN}\left[\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \Sigma \right] \end{aligned} \quad (4)$$

where the index ij represents the ith observation from the jth test. The test-specific regression coefficients β_{0j} and β_{1j} are assumed to share a common prior distribution, which is a bivariate normal distribution. Conceptually, when using eq 4 we suggest that each test has its unique features (test-specific model coefficients), but these tests are subjected to the same set of standard conditions (model coefficients are constrained to be random variables from the same prior distribution). Using the BHM in eq 4, the test-specific line defined by model coefficients β_{0j} and β_{1j} lies in between the line defined by the estimated overall mean coefficients (μ_0 and μ_1) and the line estimated using data from the individual test alone. In other words, BHM estimated test-specific models are always closer to the overall average model than the respective models fit with individual test data separately. This feature has been shown to improve overall accuracy−reduced estimation uncertainty at a group (test) level as well as improved predictive accuracy.[20]

For the ELISA test, the improved predictive accuracy is of particular interest because the fitted standard curve will be used for estimating (predicting) MC concentrations using OD measurements from water samples with unknown MC concentrations. The BHM improves the predictive accuracy of a test-

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

specific model through pooling data from multiple tests considering both unique features of individual tests and the commonality of these tests.[21] The commonality helps to correct potential bias of individual tests.[22] If we believe that the variation among tests is due to test-specific conditions, using the BHM estimated test-specific standard curve will improve the test-specific standard curve by shrinking the curve toward the overall center, thereby increasing our confidence in the resulting concentration values. If we believe that the among test variation is largely due to random noise as expected of a model with small degrees of freedom, we can use the weighted "average" model (defined by coefficients μ_0 and μ_1) for estimating MC concentrations.

In other words, regardless of the source of error we should pool data from multiple tests using BHM when estimating MC concentrations using ELISA.

MATERIALS AND METHODS

Because the uncertainty in a standard curve is not quantified, the risk of MC exposure through drinking water is uncertain. We present (1) a method for quantifying the estimation uncertainty in an estimated MC concentration using ELISA, and (2) a BHM for reducing the estimation uncertainty.

Quantifying Predictive Uncertainty. We use the Monte Carlo simulation method[19] for quantifying regression model predictive uncertainty. Briefly, the Monte Carlo simulation starts by drawing a random sample of residual variance from its sampling distribution (a rescaled χ^2 distribution), which is used to quantify the joint distribution of model coefficients (a multivariate normal distribution) for generating random samples of model coefficients. The resulting random samples of model coefficients and residual variance are then used to draw predictive samples of MC concentration.

Fitting and Evaluating BHM. We present the BHM using the log−linear model as the base model. The model is fit by using the R function `lmer` from package `lme4`.[23] Although `lmer` is an implementation of the classical linear mixed effects model, the results from `lmer` is similar to the BHM using vague or flat priors.[21] Using `lmer` simplifies the model development process.

Because the usual residual based model assessment methods are inappropriate for BHM,[20] we will not use residual-based diagnostic graphs and statistics for model assessment. Instead, we use the Watanabe-Akaike Information Criterion (WAIC),[24] a fully Bayesian implementation of information theoretical methods for assessing model predictive accuracy.[25] WAIC calculates model deviance and effective parameters based on draws from the pointwise predictive posterior distributions. This incorporates uncertainty associated with varying parameter influences (i.e., effective number of parameters), and mimics the added predictive uncertainty associated with unobserved (i.e., future) data. By comparing WAICs calculated for the BHM and the individually fit log−linear models, we demonstrate the expected improvement of model predictive accuracy of BHM. Computational details are in the Supporting Information.

Risk Assessment. With the uncertainty assessment of an estimated MC concentration value, we can use a probabilistic risk assessment approach for communicating the risk of MC exposure. Using BHM, the estimated MC concentration is expressed as a probability distribution. Based on this distribution we can express our uncertainty in terms of the probability of a concentration exceeding the criterion of 1 μg/L. State managers can then communicate the risk to the public. As an illustration, we present the estimated probability of exceeding the criterion from five water samples measured on August 1, 2014 in the Toledo Water Plant. These concentration values were estimated based on the same standard curve. Because the Toledo Water Plant did not make raw ELISA data available except for those conducted during the 2014 water crisis, we illustrate the process of pooling data from multiple tests using data from Stone Lab. In addition to the water sample that first reported a high MC concentration value that ultimately led to the "Do not drink" advisory, we also include four additional samples: a tap water sample taken on July 31 (with OD values 1.002 and 1.013), the normal control sample (0.489 and 0.470), a raw water sample taken on July 28 (0.883 and 0.933), and a diluted raw water sample taken on July 30 (0.693 and 0.645). These additional samples were selected to cover a wide rOD range. The two data points with the smallest *r*OD values are smaller than the rOD value from the smallest standard concentration. We also highlight differences in predicted MC concentrations generated from a test-specific standard curve and the two BHM standard curves (one based on β_{0j} and β_{1j} and the other based on μ_0 and μ_1 in eq 4).

Uncertainty in estimated model coefficients can be propagated into the uncertainty in the estimated MC concentration of a water sample. In our Monte Carlo simulation, we draw random samples of model coefficients. Each set of random sample forms a potential standard curve and was used to estimate a likely MC concentration based on the rOD value of a water sample. In other words, for each water sample, the Monte Carlo simulation yields the distribution of the unknown MC concentration.

RESULTS

Uncertainty Assessment using Monte Carlo Simulations. We use data from two ELISA tests conducted at the Stone Lab to illustrate the within-test uncertainty of the standard curves−one conducted on August 4, 2014 with a near perfect R^2 value (0.9993) and the other on September 11, 2014 with a respectable R^2 value of 0.96. These two R^2 values represent the highest and the lowest among the 21 assays, respectively. We also use data from tests conducted on August 1 and 2, 2014 in the Toledo water treatment plant. In all cases, we have well fit standard curves using data from these tests separately.

The Stone Lab data were transformed to fit a log−linear regression model in eq 2. Because of the small degrees of freedom, the sampling distribution of residual variance is highly variable, which translates to a considerable within test variation (Figure 1(a) and (b)). Although the "best" and the "worst" standard curves both fit to their respective data well, the predictive variations of the two models are quite different. Compared to the within test variations, the among test variability shown in the 21 individually fit standard curves (Figure 1(c)) can easily be explained as random variation.

The 4-parameter nonlinear model used by Toledo also shows substantial within and among test variations (Figure 2). Because the response variable in the 4-parameter nonlinear model is OD rather than c_{st}, the estimation uncertainty of the unknown MC concentration of a water sample is larger than it appears (Figure 2(a)). Visually, we perceive the accuracy of a model by the vertical distance between the upper and lower bounds of the 95% intervals (i.e., in the *y*-axis direction), while the uncertainty about the concentration is along the *x*-axis direction. In Figure 2(a), the dashed horizontal line represents

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

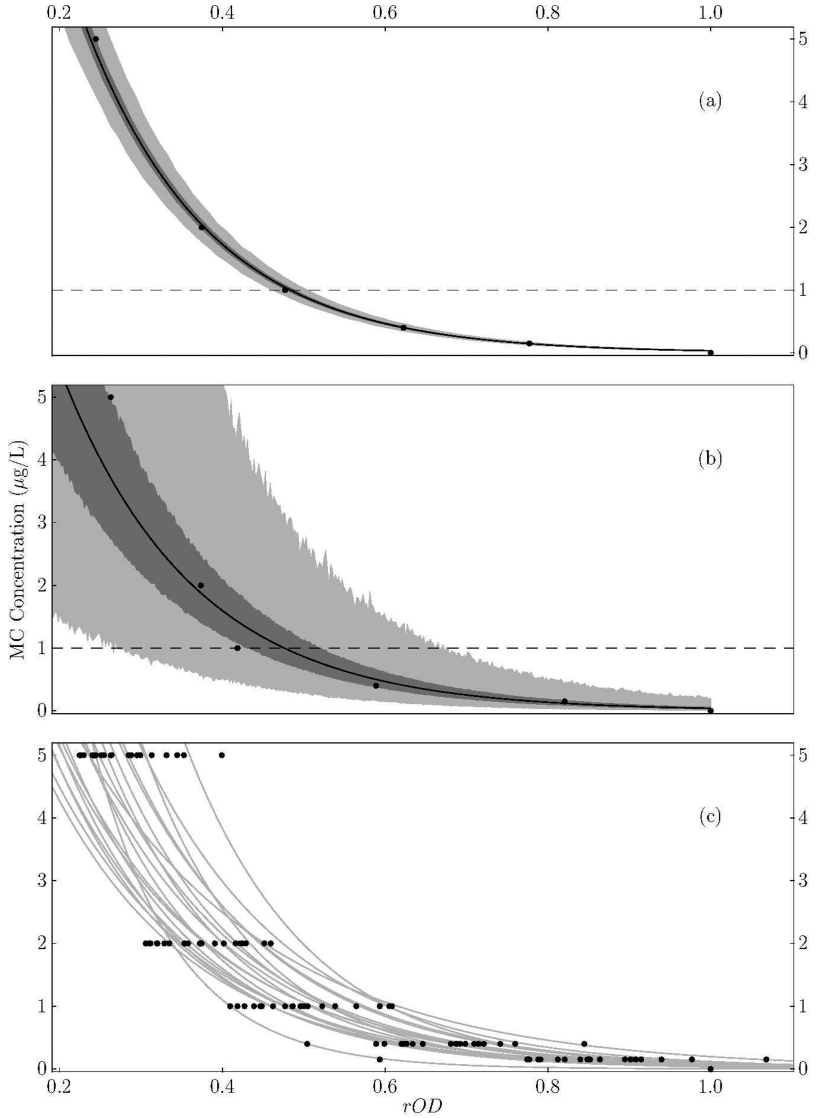


Figure 1. Uncertainty of fitted standard curves is evaluated by Monte Carlo simulation. Panel (a) shows the "best" fit standard curve (August 4, 2015), panel (b) shows the "worst" fit model (September 11, 2014), where the light-shaded area shows the 95% credible intervals of predictive distributions of MC concentrations given observed *r*OD values and the dark-shaded areas are the 50% credible intervals. Panel (c) shows the 21 individually fit standard curves.

an OD measurement from a water sample. The predicted MC concentration for this water sample is the *x*-axis value at which the horizontal dashed line intercepts the fitted standard curve. The estimation uncertainty is represented by the segment of the horizontal dashed line within the shaded area representing the predictive 95% credible interval. The estimation uncertainty in MC concentration is much larger than the figure or model statistics suggest because the regression model minimizes the prediction error in OD (the *y*-axis), not the MC concentration (the *x*-axis).

Although both models were recommended by ELISA kit manufacturer, the log−linear model is fit to minimize the estimation error of log MC concentrations, while the 4-parameter nonlinear model is fit to minimize estimation error with respect to OD. As a result, the log−linear model is better suited for estimating MC concentrations.

Bayesian Hierarchical Model. The BHM fitted test-specific models are less variable than the conventionally fit models (Figure 3). Compared to individually fit standard curves in Figure 1(c), BHM fitted test-specific standard curves are closer (both in distance, the intercept, and in shape, the slope) to the overall mean curve (black line in Figure 3).

The BHM model fit with data from all 21 tests has a WAIC of 29.8. The comparable individually fit log−linear regression model (fitted together allowing test-specific intercept and slope) has a WAIC of 37.5. Intuitively, WAIC is a measure of deviance, which is proportional to the mean squared error in our case (a normal response variable). Although the specific value of WAIC is a function of sample size (hence meaningless by itself), the difference of WAICs of two models shows the relative difference in two models predictive accuracy. The smaller WAIC of the BHM model indicates an improved predictive accuracy.

To estimate prediction uncertainty based on a test specific standard curve, we refit the standard curve to the known concentration data from the August 1, 2014 test (the Toledo data) using the log−linear model. The resulting model (the Toledo model) has an R^2 value of 0.96. The observed OD

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

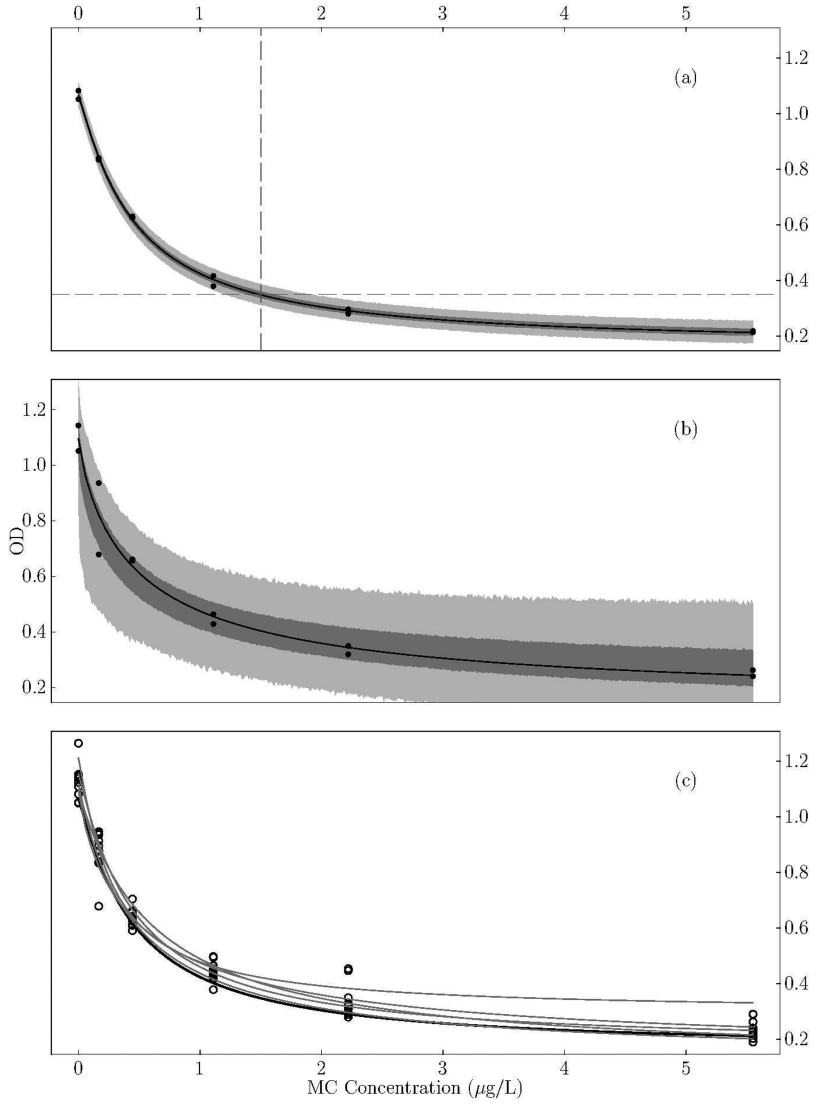


Figure 2. Uncertainty associated with the standard curve that detected the high MC concentration on August 1, 2014 in Toledo's drinking water is still considerable even with an almost perfect R^2 value (a). The within test variation of a subsequent test is, however, much larger, even though the model fits the data well (b). The six curves developed on August 1 and 2 show a large among test variation reflecting the measurement uncertainty in OD at various MC concentrations (c). The model's predictive error for MC concentration (x-axis) at a given OD value is higher (the dashed horizontal line in (a)) as the model minimizes the prediction error in OD (y-axis direction, the dashed vertical line in (a)).

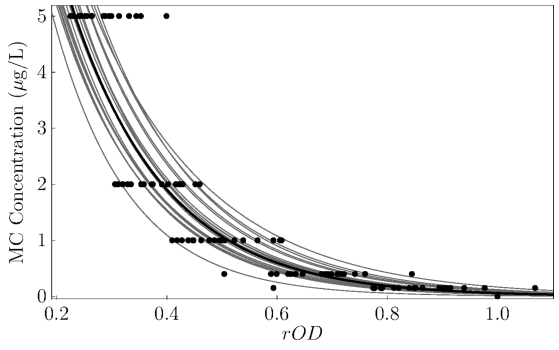


Figure 3. Standard curves fitted using BHM show a reduced among test variation (compared to Figure 1(c)). The thick dark line is the mean model characterized by μ_0 and μ_1 in eq 4. The thin shaded lines are test-specific models characterized by β_{0j} and β_{1j} where $j = 1, \cdots, 21$, the index of the 21 tests.

values for the August 1 tap water sample were 0.271 and 0.286, which translate to an average rOD value of 0.261. Using the Toledo data alone, the fitted log−linear model has an intercept of 2.48 and a slope of −5.58, resulting in an estimated MC concentration of 2.78 μg/L for this tap water sample. The uncertainty about the predicted concentration is quite large, reflected in a wide predictive 95% credible interval (CI) of [0.88, 7.70] μg/L. In other words, the resulting concentration of 2.78 is statistically not different from 1 (estimates associated with solid lines in Figure 4) based on the classical hypothesis testing approach.

Compared to the BHM developed by pooling the 21 tests from the Stone Lab (mean intercept $\mu_0 = 2.96$ and mean slope $\mu_1 = -5.80$), the single test using Toledo data produced a much lower intercept. Using the BHM model, the mean predicted MC concentration should be 4.26 μg/L (dotted lines in Figure 4). When the Toledo data are pooled as part of the Stone Lab data, the average mean model (the BHM average model) changed little ($\mu_0 = 2.94$, $\mu_1 = -5.78$), but the BHM estimated test-specific (Toledo) intercept and slope are now 2.76 and −6.03 (the Toledo BHM model), respectively (dashed

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

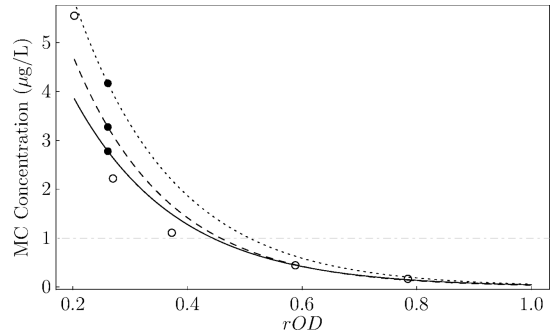


Figure 4. Predicted MC concentrations for the tap water sample collected on August 1, 2014 are estimated by three standard curves (solid circles). The solid line is the standard curve based on Toledo's test data (open circles) only. The dotted line is the mean standard curve based on the BHM (pooling data from Stone Lab and the August 1 test), and the dashed line is the BHM estimate standard curve for the Toledo test.

lines in Figure 4). The BHM estimated Toledo standard curve lies between the BHM mean curve and the curve fit with Toledo data alone (within the range of the Toledo data). Using the Toledo standard curve, the estimated mean MC concentration for the August 1 tap water sample is 3.45 μg/L with a 95% CI of (1.88, 5.91) (Figure 5).

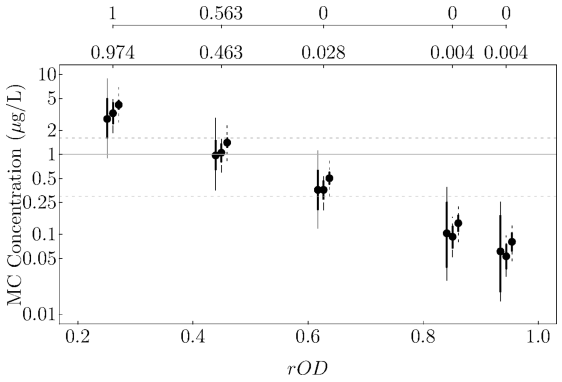


Figure 5. Estimated MC concentration distributions from the three models are shown by the 95% CI (vertical lines), where the thin lines are predictive 95% CI for individual concentration measurement and the thick lines are the 95% CI for the estimated mean concentration. The predictive distributions estimated by the Toledo model, the BHM Toledo model, and the BHM mean model (from left to right) are grouped based on the rOD values of the respective water samples. Probability of an estimated MC concentration exceeding the WHO criterion of 1 are shown on top of the figure. The numbers in the lower row are based on the Toledo model and the numbers on the upper row are based on the Toledo BHM model. The shaded horizontal lines are the three criterion values (0.3, 1, and 1.6).

Risk Assessment. In the ELISA test of August 1, 2014, the measured OD values used for developing the standard curve showed little variation (the two replicates are almost identical for all standard concentrations, Figure 2(a)), which can be an anomaly as OD observation variance can be large (Figure 2(b)) and the variability is consistent from concentration to concentration (Figure 2(c)).

We represent the estimation uncertainty in terms of a predictive distribution, which is used to estimate the probability of the underlying concentration exceeding the criterion (Figure 5). The closer the estimated probability is to 1 or 0, the more certain we are about whether the underlying concentration is above or below the criterion, respectively.

Using the standard curve fit with Toledo data alone, the estimated MC concentration distribution has a high variance (Figure 5). When we treat standard curves from multiple tests as replicates and use the BHM average model, the estimated MC concentration can be quite different from the estimate using test data alone. The Toledo BHM model estimated test-specific MC concentration lies in between these two estimates (the shrinkage effect), and the estimation uncertainty is typically reduced.

DISCUSSION

We discussed the estimation uncertainty associated with the enzyme-linked immunosorbent assay (ELISA) method for measuring microcystin (MC) concentrations. Because the estimation process depends on the standard curve fitted with a small number of known standard concentrations, uncertainty associated with the estimation from a single test is likely far larger than the R^2 value or residual variance would suggest. Using a Monte Carlo simulation, we showed that the estimation uncertainty can be substantial. Furthermore, such uncertainty is readily reflected in the among test variation—changes in the fitted standard curve from test to test. As the estimation uncertainty is rarely presented in the estimated MC concentration, decision making such as the one facing officials of Toledo can be a very difficult task.

Using the Bayesian hierarchical model (BHM) we show that estimation uncertainty can be reduced by sharing information from multiple tests. Much of the mathematical basis of BHM is discussed by Efron and Morris.[22]

Because the estimated MC concentrations are the basis for decision making in many water management problems, we see further advantages of using a Bayesian hierarchical modeling approach:

- As in any drinking water management situations, MC criterion is set to measure human exposure to the toxin. The current standard of 1 μg/L from World Health Organization is a chronic exposure standard. In other words, a harmful effect is expected when a "long-term" exposure to a concentration of 1 μg/L or above is observed. As a result, the criterion applies to the mean concentration of MC, not individual measurements. In our examples, the predictive distribution combines uncertainty in model coefficients and residual variance to estimate the uncertainty associated with individual measurements. Uncertainty about the mean is readily available (Figure 5, see Supporting Information for details.)

  U.S. Environmental Protection Agency's health advisories on algal toxins in drinking water have two microcystin concentration levels: 0.3 μg/L for children "younger than school age" and 1.6 μg/L for other ages.[26] These values are based on potential health effect from exposure for 10 days, hence the criteria should be compared to 10-day average concentrations. In any case, the current practice of comparing the estimated MC concentration from a one time sample is not appropriate. Given that most drinking water systems in Ohio monitor MC concentrations weekly or biweekly, a time-series analysis method (e.g., dynamic linear model[27]) should be used for estimating changes in the mean concentration over time.

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

- As discussed earlier, the among-test variation (changes in the fitted standard curve from test to test) is often interpreted as something associated with specific test conditions. Our simulation suggested that the among-test variation could be a result of large within test variation hidden from individual tests due to a small sample size. Using BHM provides a compromise for both possibilities. In other words, data from multiple tests should be used.

  Alternatively, the manufacturer of the test kit can carry out additional studies to provide a robust standard curve. Specifically, we can carry out the ELISA test using more than five standard concentrations. For example, we can fill the standard 96-well plate with 48 different known concentrations, each with two replicates as usual. The resulting standard curve will have a much wider inference range and a better characterization of the uncertainty. The model can be included as part of the kit and be used as a Bayesian prior model. When measuring MC concentrations, a user can carry out the test as usual. A Bayesian linear regression model[28,29] can be used to derive a standard curve that combines the prior model and test-specific data. The Bayesian method can be readily programmed into many standard software packages for routine use by lab operators. This approach is essentially the same as combining data from multiple tests, if we assume that among-test differences we routinely observed is due to observation error (no other test-specific sources of uncertainty).
- Our simulation, however, cannot definitely conclude that the test-to-test variation is due entirely to random noise. With a Bayesian hierarchical model, we have both the test-specific model and the average model. If both models predict a high probability of the mean concentration exceeding the criterion, we can conclude that the risk of harmful exposure is high.

  Individual drinking water treatment facilities in Ohio are likely to send their water samples to a small number of laboratories for testing. Data from these tests should be shared with a state regulatory agency for developing the Bayesian hierarchical model and updating the model over time, while drinking water plant operators receive the same standard reports as usual. When we have individual measurements exceeding the criterion, we should use the updated model for assessing likelihood of the mean concentration exceeding the criterion.

  Individual laboratories currently conducting ELISA testing may already have data from a large number of tests. Computer programs presented in this paper can be used by individual laboratories to develop the BHM.

Although we focused on the ELISA test for measuring MC concentrations, our results, in principle, apply to similar methods used for measure concentrations of other water quality variables such as total phosphorus and total nitrogen,[30,31] where concentration measurements are based on a standard curve developed using regression with a few degrees of freedom.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b03029.

Details of analysis, WAIC for model comparisons, R code, link to source data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: song.qian@utoledo.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Much of the work reported in this paper was carried out in a special topics class on data analysis related to quantifying microcystin concentrations taught in fall 2014. We thank Tim Fisher, chair of the Department of Environmental Sciences of the University of Toledo, for approving the class in the wake of the Toledo drinking water crisis, City of Toledo for sharing ELISA test data measured during the Toledo water crisis, Briana Zellner for conducting many of the ELISAs at Stone Lab, and the Ohio EPA for funding the ELISA tests and coordination of sample delivery to Stone Lab. Comments and recommendations from Heather Raymond on an early version of the paper are greatly appreciated. S. Qian's work was supported in part by U.S. EPA (grant GL-00E01284) and NOAA (through the Cooperative Institute for Limnology and Ecosystems Research of the University of Michigan at Ann Arbor). We thank the associate editor and three reviewers, whose comments and suggestions led to a greatly improved paper.

■ REFERENCES

(1) Smith, V. Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environ. Sci. Pollut. Res.* **2003**, *10*, 126−139.

(2) Paerl, H.; Huisman, J. Blooms Like It Hot. *Science* **2008**, *320*, 57−58.

(3) DeMott, W.; Zhang, Q.-X.; Carmichael, W. Effects of Toxic Cyanobacteria and Purified Toxins on the Survival and Feeding of a Copepod and Three Species of Daphnia. *Limnol. Oceanogr.* **1991**, *36*, 1346−1357.

(4) Carmichael, W.; Azevedo, S.; An, J.; Molica, R.; Jochimsen, E.; Lau, S.; Rinehart, K.; Shaw, G.; Eaglesham, G. Human fatalities from cyanobacteria: Chemical and biological evidence for cyanotoxins. *Environ. Health Perspect.* **2001**, *109*, 663−668.

(5) *Harmful Cyanobacteria*, Aquatic Ecology Series; Huisman, J., Matthijs, H., Visser, P., Eds.; Springer: Netherlands, 2005; Vol. 3.

(6) Ibelings, B.; Backer, L.; Kardinaal, W.; Chorus, I. Current approaches to cyanotoxin risk assessment and risk management around the globe. *Harmful Algae* **2014**, *40*, 63−74.

(7) World Health Organization. *Guidelines for Drinking-water Quality. 2nd ed., addendum to Volume 2, Health Criteria and Other Supporting Information*; WHO: Geneva, 1998. http://www.who.int/water_sanitation_health/dwq/gdwq2v1/en/index4.html.

(8) Thurber, M. The Toledo water crisis that wasn't.http://watchdog.org/165104/toledo-water-crisis/.

(9) Bridgeman, T.; Chaffin, J.; Filbrun, J. A novel method for tracking western Lake Erie Microcystis blooms, 2002−2011. *J. Great Lakes Res.* **2013**, *39*, 83−89.

(10) Rinta-Kanto, J. M.; Wilhelm, S. W. Diversity of microcystin-producing cyanobacteria in spatially isolated regions of Lake Erie. *Applied and environmental microbiology* **2006**, *72*, 5083−5085.

(11) Stumpf, R.; Wynne, T.; Baker, D.; Fahnenstiel, G. Interannual Variability of Cyanobacterial Blooms in Lake Erie. *PLoS One* **2012**, *7*, e42444.

(12) Michalak, A.; et al. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

expected future conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6448−6452.
(13) Carmichael, W. Cyanobacteria secondary metabolites−the cyanotoxins. *J. Appl. Bacteriol.* **1992**, *72*, 445−459.
(14) Chu, F.; Huang, X.; Wei, R. Enzyme-linked immunosorbent assay for microcystins in blue-green algal blooms. *Journal of Association of Official Analytical Chemists* **1990**, *73*, 451−456.
(15) Fischer, W.; Garthwaite, I.; Miles, C.; Ross, K.; Aggen, J.; Chamberlin, A.; Towers, N.; Dietrich, D. Congener-independent immunoassay for microcystins and nodularins. *Environ. Sci. Technol.* **2001**, *35*, 4849−4856.
(16) Weisberg, S. *Applied Linear Regression*; Wiley, 2005.
(17) Sprugel, D. G. Correcting for bias in log-transformed allometric equations. *Ecology* **1983**, *64*, 209−210.
(18) Stow, C.; Reckhow, K.; Qian, S. A Bayesian approach to retransformation bias in transformed regression. *Ecology* **2006**, *87*, 1472−1477.
(19) Qian, S. *Environmental and Ecological Statistics with R*; Chapman and Hall/CRC Press, 2010.
(20) Qian, S.; Stow, C.; Cha, Y. Implications of Stein's Paradox for Environmental Standard Compliance Assessment. *Environ. Sci. Technol.* **2015**, *49*, 5913−5920.
(21) Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Cambridge University Press, New York, 2007.
(22) Efron, B.; Morris, C. Stein's Paradox in Statistics. *Sci. Am.* **1977**, *236*, 119−127.
(23) Bates, D. *lme4: Mixed-Effects Modeling with R*; Springer, 2010.
(24) Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **2010**, *11*, 3571−3594.
(25) Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **2014**, *24*, 997−1016.
(26) EPA Newsroom, EPA Issues Health Advisories to Protect Americans from Algal Toxins in Drinking Water. 2015; http://yosemite.epa.gov/opa/admpress.nsf/3881d73f4d4aaa0b85257359003f5348/547dc50c15c82aaf85257e3d004d7f67!opendocument.
(27) Petris, G.; Petrone, S.; Campagnoli, P. *Dynamic Linear Models with R*; Springer, 2009.
(28) Lindley, D.; Smith, A. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, B* **1972**, *34* (1), 1−41.
(29) Box, G.; Tiao, G. *Bayesian Inference in Statistical Analysis*; Addison-Wesley, Reading, MA, 1973.
(30) APHA. *Standard Methods for the Examination of Water and Wastewater*, 22nd ed.; American Public Health Association, 2012.
(31) Wetzel, R.; Likens, G. *Limnological Analysis*, 3rd ed.; Springer: New York, 2000.

DOI: 10.1021/acs.est.5b03029
Environ. Sci. Technol. XXXX, XXX, XXX−XXX