Rethinking Calibration as A Statistical Estimation Problem to Improve Measurement Accuracy

Sabrina Jaffe[1†], Duane Gossiaux[2], Reagan M. Errera[2], Emanuela Gionfriddo[3], Hongjun Wang[4], Curtis J. Richardson[4], Nipunika H. Godage[3] and Song S. Qian[1*†]

[1]Department of Environmental Sciences, The University of Toledo, 2802 Bancroft Street, Toledo, 43606, Ohio, USA.
[2]Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, 4840 South State Road, Ann Arbor, 48108, Michigan, USA.
[3]Department of Chemistry, The University of Toledo, 2802 Bancroft Street, Toledo, 43606, Ohio, USA.
[4]Nicholas School of the Environment, Duke University, 9 Circuit Drive, Durham, 27710, North Carolina, USA.

*Corresponding author(s). E-mail(s): song.qian@utoledo.edu;
Contributing authors: sabrina.jaffe@rockets.utoledo.edu;
duane.gossiaux@noaa.gov; reagan.errera@noaa.gov;
emanuela.giofriddo@utoledo.edu; hjwang78@gmail.com;
curtr@duke.edu; nhirimu@rocketes.utoledo.edu;
†These authors contributed equally to this work.

Abstract

Purpose: In this study, we reevaluated the commonly used calibration-based measurement method from the perspective of statistical estimation. Our aim was to highlight the substantial uncertainty inherent in the resulting measurements. To address this issue, we proposed a Bayesian hierarchical modeling (BHM) approach, which offers enhanced accuracy and consistency for calibration-based methods.
Methods: Utilizing Monte Carlo simulation, we demonstrated the considerable variability observed in calibration measurements and illustrated the improvement achieved through the implementation of the BHM approach.
Results: By analyzing data obtained from a long-term water quality monitoring program, we identified that the notable variability in a typical calibration-based method arises due to the relatively limited sample size used for fitting the calibration curve. The BHM approach effectively mitigated this uncertainty by pooling relevant information from multiple data points within a test and combining information from calibration curve coefficients across similar calibration curves. A novel computation method facilitated the seamless integration of BHM without necessitating alterations to existing laboratory procedures.
Conclusions: Our findings demonstrate that the accuracy of all calibration-based measurement methods can be significantly enhanced by replacing the conventional regression method with the more robust BHM approach. The incorporation of Bayesian hierarchical modeling provides a promising avenue for improving the reliability and precision of calibration-based methods, addressing the challenges posed by high measurement variability.

Keywords: Bayesian statistics, Calibration, hierarchical modeling, missing data problem, sequential updating

1 Introduction

Calibration-curve based methods are extensively used in analytical chemistry to determine substance (analyte) concentrations, constituting over 90% of chemical analytical work [1]. This process typically involves two distinct steps: (1) developing a calibration or standard curve through empirical modeling, using known concentrations of standard solutions as predictors and corresponding instrument-generated responses as the response variable, and (2) estimating the unknown analyte concentration of a sample based on the

inverse function of the resulting regression model using its instrumental response. Although calibration standard solutions and samples with unknown concentrations are often measured simultaneously, these two steps are independent. The accuracy of estimated unknown analyte concentrations depends on the statistical characteristics of the regression model.

Typically, calibration processes rely on goodness-of-fit statistics (e.g., coefficient of variation or R^2-value for linear regression) as a summary of the regression model's "quality," while predictive accuracy statistics are rarely reported. As these statistics measure the fit and prediction of the instrumental responses (response variable), they are not directly relevant to the classical calibration problem, which aims to estimate unknown analyte concentrations (the missing predictor variable values in the combined data from both standard solutions and calibration samples). To evaluate the uncertainty of the estimate, statistical characteristics (the sampling distribution) of the classical calibration estimator (inverse function of the regression model) must be derived. The sampling distribution of an estimator is the statistical measure of estimation uncertainty, often expressed as the confidence interval of an estimate. Sampling distribution for the calibration estimate is only available for the linear calibration curve, a Cauchy distribution for which the mean and variance do not exist [2]. As such, quantifying the uncertainty of the estimated concentration is contentious and almost never reported in practice [2, 3]. Nevertheless, the calibration estimator is often highly variable, primarily due to the limited number of standard solutions used in typical calibration methods limited by instrument capacity [4, 5].

The variability resulting from the high level of uncertainty in the classical calibration estimator leads to significant fluctuations in "measured" concentrations, despite the lack of reporting estimation uncertainty. To

improve the accuracy of calibration methods, it is essential to determine the extent of uncertainty and identify ways to reduce it. The difficulty of estimating quantifying uncertainty in the classical calibration problem can be circumvented by using a Bayesian method for estimating the posterior distribution of unknown analyte concentrations. Mathematically, the common inference basis of the two approaches is the likelihood function. By using vague or flat priors, the joint posterior distribution of model parameters and unknown analyte concentrations becomes the normalized likelihood function, enabling the derivation of marginal posterior distributions for summarizing estimation uncertainty [6]. Modern simulation-based Bayesian computation method (i.e., Markov chain Monte Carlo simulation) allows us to easily derive these posterior distributions.

However, using a flat prior does not change the unbiased nature of the classical calibration estimator, which often inflates the variance of the estimator. Statisticians have long learned that "certain deliberately induced biases can drastically improve estimation properties when there are several parameters to be estimated" similtaneously [7]. By employing the Bayesian method with proper informative priors, biases can be introduced to reduce estimation uncertainty by incorporating relevant information beyond the available data [8]. In the context of a calibration problem, two pertinent sources of information can be utilized to diminish estimation (or measurement) uncertainty.

Firstly, we typically process multiple calibration samples with unknown analyte concentrations that need to be estimated simultaneously. The situation aligns with Stein's paradox [8], which suggests that biased estimators, such as the classical James-Stein estimator [9, 10] and empirical Bayes methods [11], can improve the overall estimation accuracy when

estimating more than three variables simultaneously. Secondly, laboratories conducting routine analyses accumulate multiple calibration curves over time. Through the use of Bayesian hierarchical modeling (BHM) methods, we can achieve a similar improvement in estimation accuracy for the calibration curve coefficients [12]. In both cases, the approach involves shrinking individually estimated (and unbiased) quantities (multiple unknown concentrations within a calibration test and calibration curve coefficients among multiple tests) towards their corresponding overall averages. This mathematical technique is proven to reduce overall estimation uncertainty.

The reason of this improvement in estimation accuracy can be understood intuitively as follows: when we empirically estimate a quantity (with inherent error), the estimate may be either overestimated or underestimated compared to the underlying true quantity. In isolation, we are unsure of the direction of the estimation error, making an unbiased estimator desirable. However, when there are multiple estimates of the same quantity from similar studies, the overall average of these estimates serves as a reference to determine whether an estimate is likely too high or too low. Consequently, shrinking the estimates towards their overall mean is advantageous in reducing overall estimation uncertainty, as demonstrated by mathematical proof [8]. The James-Stein estimator and BHM calculate the levels of shrinkage based on the strength of the information derived from the relative magnitudes of the estimation uncertainty of individual estimates compared to the spread of individual estimates.

In this paper, we demonstrate how the classical calibration method results in substantial estimation uncertainty and show how this uncertainty can be effectively reduced by adopting the BHM approach. Our strategy involves combining the calibration curve fitting and unknown concentration

estimation into a single Bayesian parameter estimation problem. To address the computational challenges associated with implementing the BHM approach in a typical analytical laboratory setting for calibration problems, particularly when leveraging information from multiple samples and curves, we have devised a computer algorithm that systematically accumulates information from existing calibration curves of the same substance, thereby diminishing estimation uncertainty without necesitating any modifications to the current lab procedures.

To assess the practical feasibility of our BHM approach, we applied it to data from a long-term water quality monitoring program conducted by the Great Lakes Environmental Research Laboratory (GLERL) of the National Oceanography and Atmospheric Administration (NOAA). This program has been regularly monitoring cyanobacterial toxin microcystin (MC) concentrations in the western basin of Lake Erie since 2012. Microcystins are a class of cyclic heptapeptides representing over 240 identified compounds [13–15] with varying toxicities. Toxic blooms of Microcystis spp., the major freshwater cyanobacteria producer of microcystins, occur annually within the western basin of Lake Erie [16] impacting both human and ecosystem health. The cyanobacteria harmful algal bloom (HAB) is extensively monitored to provide stakeholders (water managers, public, etc.) with data regarding bloom toxin concentrations. GLERL established eight monitoring sites in 2012 that are sampled weekly during the HAB season for a variety of water quality parameters, included microcystin concentrations [16]. Microcystins concentrations are analyzed using an ELISA kit (Abraxis) and data are distributed to stakeholders within 48-hours of collection.

Through our study, we illustrate how the BHM approach can be effectively implemented in real-world laboratory settings and how it

significantly improves estimation accuracy when estimating unknown concentrations using the calibration method.

2 Results

We obtained 214 sets of enzyme-linked immunosorbent assay (ELISA) test results from NOAA-GLERL, encompassing data processed between 2012 and 2021. The ELISA tests utilized commercial test kits from Abraxis, Inc., comprising six standard solutions with known MC concentrations ranging from 0 to 5.00 μg/L. Additionally, each test involved a control sample with an MC concentration of 0.75 μg/L. Up to 40 water samples with unknown MC concentrations could be tested with each test kit, and all samples were tested with a replicate. Before fitting the calibration curve and estimating unknown MC concentrations, the measured instrumental responses from each pair of replicates were averaged.

The instrumental responses were fit as a linear function of log MC concentrations for tests conducted until 18 July 2016 (tests 1-84). The calibration curve followed the form $\log(y) = \beta_0 + \beta_1 \log(x) + \varepsilon$, where y represents the instrumental response, x is the known MC concentration, and ε is a random variable following a normal distribution $N(0, \sigma^2)$). Due to the log transformation of x, the calibration coefficients (β_0 and β_1) are estimated based on the five non-zero standard solutions, resulting in a regression model fit with five data points. Consequently, the residual variance σ^2 was estimated with degrees of freedom of 3.

Starting from 25 July 2016 (tests 85-214), the calibration curve was described by a nonlinear regression model in the form of the four-parameter logistic function as shown in equation (1). This model involves four unknown parameters and has a degree of freedom of 1. The small degrees of freedom

leads to increased variability in the resulting regression model. In both the linear and nonlinear calibration models, the degrees of freedom are below 4, making it impossible to perform a reliable statistical assessment of predictive uncertainty for the fitted regression models.

$$y = \theta_4 + \frac{\theta_1 - \theta_4}{1 + \left(\frac{x}{\theta_3}\right)^{\theta_2}} + \varepsilon. \tag{1}$$

In our analysis, we employed the nonlinear calibration curve (equation 1) for all 214 tests to estimate MC concentrations. We used six different modeling methods for model evaluation:

1. Standard inverse-function estimator (BC5): Using Abaxis ELISA test kits, we fitted calibration curves fitted with five observations from the standard solutions.
2. Inverse-Function Estimator with All 12 Standard Sample Observations (BC12): The calibration curve was fit using all 12 standard sample observations. The inverse function of the calibration curves of BC5 and BC12 were used to estimate the concentration of the quality control sample.
3. Bayesian Estimator (Bayes): This method combines the fitting and estimating processes, without leveraging information within and across tests (i.e., using relatively vague priors for all unknown quantities).
4. BHM with Information Shared Within Each Test (BHM1): Information from all unknown calibration samples within a test is shared to improve estimation accuracy.
5. BHM with Information Shared Across Each Test (BHM2): Information of calibration curve coefficients across all tests is shared to reduce estimation uncertainty.
6. BHM with Information Shared Within and Across All Tests (BHM3).

Additional details concerning these six methods can be found in the Methods section and supporting materials.

We compare the results of BC5 and BC12 to assess the impact of sample size on calibration results. The outcome of BC12 is then contrasted with the four Bayesian methods (Bayes, BHM1, BHM2, and BHM3), which used all 12 standard sample observations and instrumental responses from water samples with unknown MC concentrations to fit the standard curve. For BHM2 and BHM3, we implemented a sequential updating algorithm, where tests 1-9 were used to establish the initial priors. The algorithm allows us to evaluate tests 10-214 one at a time.

Throughout the 214 tests, the instrumental responses exhibited considerable variation (Figure 1).

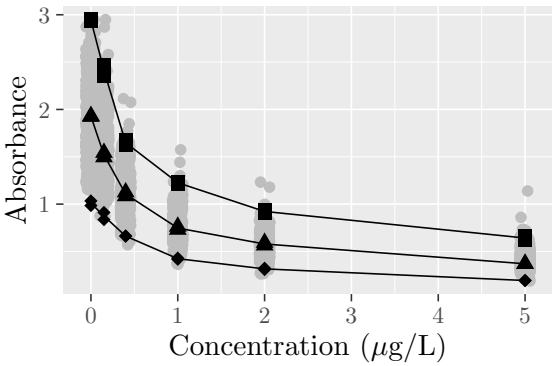

Fig. 1 Raw ELISA testing data from GLERL Western Lake Erie Harmful Algal Bloom Monitoring Program. The shaded solid circles are the measured standard solution data points from the 214 tests. The black shapes are the results from three tests with the highest (squares) lowest (diamonds) and median response value (triangles) at the 0 concentration point.

In our analysis, we used posterior simulation for BC5 and BC12 to estimate their estimation uncertainty (as illustrated in Chapter 9 of [5]). Within each Abraxis kit, a control sample with a known MC concentration of 0.75 μg/L was provided. To assess estimation accuracy, we compared the

posterior distribution of the estimated control sample MC concentration (represented by 5,000 random samples from Monte Carlo simulations) with the known value.

To quantify the estimation uncertainty for each test, we calculated absolute values of the differences between the 5,000 random samples and 0.75 μg/L the known value of the control sample. The median of these absolute differences represents the deviance of the estimated values from the true value, which is commonly referred to as the bias (Figure 2).

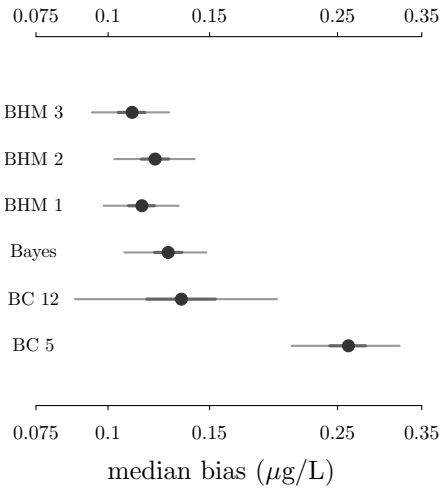

Fig. 2 Estimation bias of the quality control sample from the six methods (listed on the vertical axis) are compared. The bias is the absolute value of the difference betwen the estimated and the true concentration value (0.75 μg/L)), labeled on the horizontal axis. The solid circles are the medians, the black (thick) lines are the 50% credible intervals, the gray (thin) lines are the 95% credible intervals.

As expected, increasing the sample size for fitting the calibration curve from 5 (degrees of freedom or df=1) to 12 (df=8) significantly improved the estimation accuracy. This reduction is evidenced by median biases of 0.262

and 0.134 for BC5 and BC12, respectively. The bias is in comparison to the known concentration of 0.75 μg/L).

When employing the Bayesian estimator (Bayes), the median bias is 0.127 μg/L. This value is smaller than that of the inverse function estimator (BC12), a result of using the weakly informative prior [17] on all unknown concentrations, which prevented extreme values of the estimated concentrations. As such, we expect that the Bayes estimator will consistently outperform the inverse-function estimator.

By implementing the BHM approach to utilize relevant information, log-biases were further reduced. Leveraging multiple water samples within a test only (BHM1) resulted in a median bias of 0.11 μg/L, while leveraging multiple calibration curves to improve estimation accuracy of calibration curve coefficients only (BHM2) resulted in median log bias of 0.12. The most effective method was leveraging both within and across tests (BHM3), with the smallest bias of 0.11.

The four Bayesian estimators produced log bias distributions with comparable and low variances, indicating improved consistency. In contrast, the two inverse estimators have much larger variances in log bias distributions (Figure 2). We note that the spread of BC5 log bias is considerably underestimated due to a large number of posterior simulations of the BC5 model resulting in non-real estimates. Likewise, the log bias of BC12 is also underestimated to a lesser extent (see supporting materials for details). In short, Bayesian estimators demonstrate smaller bias and greater consistency compared to the inverse-function estimators.

The results in Figure 2 illustrates the trend of reducing estimation uncertainty as more relevant information is incorporated in the estimation process. In this particular example, pooling unknown concentrations of

calibration samples within a test appears to have a greater effect than pooling information on calibration coefficients across multiple calibration curves. This is primarily due to the large difference among the multiple curves (Figure 1), diluting the relevance of cross-curve referencing. In the supporting materials, we present additional results that highlight the significant improvement in measurement accuracy achieved through the use of the BHM method both within and across tests. These examples include:

- The Toledo Water Crisis: We provide a detailed illustration of Bayesian computation using data from six ELISA tests conducted in 2014 during the Toledo water crisis. This event occurred when an unusually high MC concentration was detected in the treated drinking water from the City of Toledo, Ohio, leading to a "do-not-drink" order for nearly half a million residents for three days.
- Linear Calibration Example: We offer an example of a calibration method for estimating orthophosphate concentrations. This method is widely regarded as highly stable and accurate.
- Determining Matrix Effects in Monitoring Xenobiotics: Another application of BHM is showcased in this example, where we investigate matrix effects in monitoring xenobiotics in the blood plasma of different mammals (as reported in [18]). The data is obtained using solid phase microextraction coupled to liquid chromatography-mass spectrometry. Through the BHM method, we assess the impact of various matrices on measurement results.

3 Methods

3.1 Calibration as a Missing Data Estimation Problem

The classical calibration problem involves a statistical estimation process that aims to estimate both the calibration curve coefficients and the unknown analyte concentrations. In this processes, we measure the instrumental responses of I standard solutions $(y_i, i = 1, \cdots, I)$ and J calibration samples with unknown analyte concentration $(y_0^j, 1, \cdots, J)$ together. The goal is to simultaneously estimate the calibration curve coefficients (θ) and the missing concentration values (x_0) together. This can be formulated as a regression problem with missing data, expressed by equation (2):

$$\begin{aligned} y_i &= f(x_i, \theta) + \varepsilon_i \\ y_0^j &= f(x_0^j, \theta) + \varepsilon_j \\ \varepsilon_{ij} &= N(0, \sigma^2) \end{aligned} \tag{2}$$

Here, f represents the calibration model (e.g., the four-parameter logistic function of equation 1). The model error term ε is assumed to follow a normal distribution with mean 0, and variance σ^2).

The joint likelihood function of the unknown parameters θ, σ^2 and x_0^j is:

$$L = \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - f(x_i,\theta))^2}{2\sigma^2}} \times \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(y_0^j - f(x_0^j,\theta))^2}{2\sigma^2}}, \tag{3}$$

where I and J are the numbers of standard solutions and calibration samples, respectively. The maximum likelihood estimator coincides with the two-step process of fitting the regression model to data from standard solutions and estimating calibration sample concentrations using the inverse

function of the fitted calibration curve. Mathematically, this is expressed as:

$$\hat{x}_0^j = f^{-1}(y_0^j, \hat{\theta}). \tag{4}$$

Here, $\hat{x}_0^j$ represents the estimated concentration of the jth calibration sample. To better quantify the uncertainty of the estimated concentrations (x_0^j), we use the Bayesian method.

3.2 Bayesian Calibration

Under the Bayesian framework, we estimate the posterior distribution of all unknown quantities, which is

$$\pi(\theta, \sigma^2, x_0^j \mid \text{data}) \propto \pi(\theta, \sigma^2, x_0^j) \times L \tag{5}$$

where $\pi(\cdot)$ represents a probability distribution function and L is the likelihood function (equation (3)).

When using non-informative (or flat) priors (i.e., $\pi(\theta, \sigma^2, x_0^j) \propto 1$), the Bayesian posterior distribution is essentially the normalized likelihood function. (For computational stability, we used the default weakly informative prior [17] implemented in the software.) Although an analytic solution of equation (5) can be challenging even for the simplest linear calibration problem [19], the Markov chain Monte Carlo (MCMC) simulation-based computation method can be readily used for estimating the marginal posterior distributions of x_0^j, θ, and σ^2.

We implement the MCMC method using Stan [20] in R [21] through the R package `rstan` [22]. The Bayesian estimation method enables us to

accurately estimate the estimation uncertainty by generating random samples of x_0^j from its marginal posterior distribution $\pi(x_0^j \mid y_0^j, y_i, x_i)$.

3.3 Bayesian Hierarchical Modeling

Because of the presence of multiple calibration samples with unknown analyte concentrations and multiple calibration tests over time, we use a Bayesian hierarchical modeling approach to reduce estimation uncertainty at two levels.

Within a single test, we impose the same prior distribution for all unknown concentrations to reflect our knowledge that these concentrations are likely different from each other, while acknowledging our ignorance of their relative magnitude:

$$\log(x_0^j) \sim N\left(\mu_{x_0}, \sigma_{x_0}^2\right) \tag{6}$$

The log transformation of x_0 is used because concentration variables can usually be approximated by the log-normal distribution [23]. The prior distribution $N\left(\mu_{x_0}, \sigma_{x_0}^2\right)$ is known as the hyper-distribution and its parameters (μ_{x_0} and $\sigma_{x_0}^2$) are known as hyper-parameters. We used weakly informative priors for hyper-parameters when fitting a hierarchical model.

Furthermore, in cases where labs routinely repeat the same test, resulting in multiple sets of estimated calibration curve coefficients, we can use another Bayesian hierarchical framework to unite these multiple tests. This approach allows us to achieve "deliberately induced biases" that enhance estimation accuracy [7]. As calibration curve coefficients are regression coefficients, we can impose the same normal prior distribution to reflect our knowledge that model coefficients vary from test to test, but we cannot

foreseen the relative magnitude of them among multiple tests:

$$\theta_k \sim N(\mu_\theta, \sigma_\theta^2). \tag{7}$$

Similarly, weakly informative priors can be used for μ_θ and σ_θ.

3.4 Sequential Updating

The BHM approach is ideal for reducing the estimation uncertainty in individual test. However, running a BHM model can be impractical because it requires the availability of data from multiple tests. Even for labs with such data already available, using BHM for each additional test requires combining the most recent data with data from previous tests. As the number of tests increases, not only is the process of combining data cumbersome, but also the computational burden will inevitably become increasingly intolerable.

Calibration curve coefficients for a specific calibration curve are represented as random variables with a normal distribution (equation (7)). We can interpret the hyper-distribution ($N(\mu_\theta, \sigma_\theta^2)$) as the common prior distribution for all tests. Once we have fitted the BHM with data from a large number of tests, we can consider the estimated posterior μ_θ and σ_θ^2 as the prior mean and variance for coefficient of a future test [11].

The conditional posterior distribution of $\mu_\theta \mid \sigma_\theta^2$ can be shown to be a normal distribution and the posterior distribution of σ_θ^2 can be quantified as proportional to a scaled inverse gamma (IG) distribution [24]. Accordingly, we can approximate the joint posterior distribution of μ_θ and σ_θ^2 using a

normal–inverse-gamma (N-IG) distribution with four parameters [25]:

$$\begin{aligned} \mu_\theta \mid \sigma_\theta &\sim N(\mu_0, \sigma_\theta/\lambda) \\ \sigma_\theta &\sim IG(\alpha, \beta) \end{aligned}.$$

This distribution can then be used as the prior distribution for the next calibration curve coefficient. Consequently, we can fit the next calibration curve using the Bayesian method with informative priors for model coefficients derived from the posterior from the previous BHM model, thereby avoiding fitting the BHM combining data from all tests.

Since we use Markov chain Monte Carlo method, the posterior distribution of μ_θ and σ_θ^2 are represented by random samples, we can derive the four distribution parameters ($\mu_0, \lambda, \alpha,$ and β) using the method of moments. That is, given random samples of μ_θ and σ_θ^2, we calculate their sample means and variances, and equate them to the theoretical mean and variance formulae of the two parameters:

$$\begin{aligned} E(\theta) &= \mu, & Var(\theta) &= \tfrac{\beta_\theta}{(\alpha_\theta - 1)\lambda_\theta} \\ E(\sigma_\theta^2) &= \tfrac{\beta_\theta}{\alpha_\theta - 1}, & Var(\sigma_\theta^2) &= \tfrac{\beta_\theta^2}{(\alpha_\theta - 1)^2(\alpha_\theta - 2)} \end{aligned}$$

Solving for the unknowns parameters:

$$\begin{aligned} \mu_\theta^0 &= E(\theta), & \lambda_\theta &= E(\sigma_\theta^2)/Var(\theta) \\ \alpha_\theta &= 2 + E^2(\sigma_\theta^2)/Var(\sigma_\theta^2), & \beta_\theta &= E(\sigma_\theta^2)(\alpha_\theta - 1) \end{aligned}$$

The joint prior distribution for the hyper-parameters is specified by four hyper-parameters: $\mu_0, \lambda, \alpha,$ and β. We can start with relatively flat prior (e.g., $\mu_0 = 0, \lambda = 1, \alpha = 0.01,$ and $\beta = 0.01$). Assuming the joint posterior distribution of μ_θ and σ_θ^2 can be represented by a N-IG distribution, the posterior parameters can be summarized using the method of moments.

These parameters are then used as prior parameters for analyzing data from the next test and their updated posterior distribution can be used to derive the prior for the next iteration. After a number of rounds of updating, the posterior parameters should converge and the subsequent updatings are essentially fitting a Bayesian linear/nonlinear regression model. Given that most labs conducting ELISA tests have data from previous tests, implementing sequential updating process is feasible.

4 Discussion

We identified an inherent statistical weakness of the calibration-based chemical measurement methods, that leads to high levels of uncertainty in resulting measurements. As these methods are used for quantifying concentrations of nearly all substances and affecting all aspects of our daily life, the importance of our study cannot be overstated. Our study discussed that the underlying statistical reason for the method's high level of estimation uncertainty is the small sample size (usually limited to 6 standard solutions) used in fitting the calibration curve.

Because the calibration method has been used by nearly all chemical measurement processes for a long time, existing lab procedures, equipment, and related infrastructure are fully adapted to the current standards. Consequently, requiring a larger sample size to fit the calibration curve may be costly and impractical. Our study recommends the use of Bayesian hierarchical modeling (BHM) approach to improve measurement accuracy by leveraging relevant information already available in a typical chemical laboratory.

Although the BHM approach presented in this paper is not new (both the frequentist and Bayesian statistical theories showed the benefit of sharing

information among relevant sources in improving estimation accuracy [9, 10, 26]), our contributions lie in three areas. First, we illustrated the accuracy problem of calibration methods, a problem which is largely ignored in the field as estimation uncertainty is almost never reported in practice. Second, we demonstrated the effectiveness of the BHM approach in improving the accuracy. This approach is easily applicable due to our third contribution: the introduction of the sequential updating algorithm, which can be automated by incorporating it into a computer application, thereby replacing the currently used applications, often based on spreadsheets.

We envision that a web-browser-based application (e.g., one based on the R Shiny package [27]) would be developed to incorporate our sequential updating algorithm for each laboratory to replace the currently used classical calibration estimator, primarily implemented in custom spreadsheet-based software. Such a browser-based app would necessarily be lab- and analyte-specific, as is the currently used spreadsheet-based software, to accumulate the updated hyper-distributions of calibration curve coefficients.

Acknowledgments. Qian and Jaffe's work is partially supported by a grant from Ohio Sea Grant (2020-2023), Qian and Gionfriddo were partially supported by the University of Toledo Interdiscuplanry grant (2020-2021).

Declarations

- Authors declare no conflict of interest/competing interests
- Authors' contributions:
  - SSQ – Conceptualization, study and data analysis design, drafting the manuscript and the supporting materials, securing funding
  - SJ – Study design, data analysis, drafting/editing manuscript

- EG – Study design, data curation, securing funding, reviewing/editing manuscript/supporting materials
- NHG – data analysis, data curation
- DG – Data curation, reviewing manuscript
- RME – reviewing/editing manuscript, data curation
- HW and CJR – reviewing/editing manuscript/supporting materials, data curation

Data Availability

Data and computer code are available at the corresponding author's GitHub repository (GitHub.com/songsqian/calibration)

References

[1] Miller, J.N., Miller, J.C.: Statistics and Chemometrics for Analytical Chemistry, 6th edn. Pearson/Prentice Hall, Essex, UK (2010)

[2] Osborne, C.: Statistical calibration: A review. International Statistical Review **59**(3), 309–336 (1991)

[3] Mandel, J.: A note on confidence intervals in regression problems **29**(3), 903–907 (1958)

[4] Qian, S.S., Chaffin, J.D., DuFour, M.R., Sherman, J.J., Golnick, P.C., Collier, C.D., Nummer, S.A., Margida, M.G.: Quantifying and reducing uncertainty in estimated microcystin concentrations from the ELISA method. Environmental Science and Technology **49**(24), 14221–14229 (2015)

[5] Qian, S.S.: Environmental and Ecological Statistics with R, 2nd edn. Chapman and Hall/CRC Press, Boca Raton, Florida (2016)

[6] Little, R.J.A., Rubin, D.B.: Statistical Analysis with Missing Data, 3rd Edition. Wiley, New York (2019)

[7] Efron, B.: Biased versus unbiased estimation. Advances in Mathematics **16**, 259–277 (1975)

[8] Efron, B., Morris, C.: Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association **70**(350), 311–319 (1975)

[9] Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 197–206. University of California Press, Berkeley, California (1956)

[10] James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability vol. 1, pp. 361–379. University of California Press, Berkeley, California (1961)

[11] Efron, B.: Empirical Bayes methods for combining likelihoods. Journal of the American Statistical Association **91**(434), 538–550 (1996)

[12] Gelman, A., Chew, G.L., Shnaidman, M.: Bayesian analysis of serial dilution assays. Biometrics **60**, 407–417 (2004)

[13] Svircev, Z., Drobac, D., Tokodi, N., Mijovic, B., Codd, G.A., Meriluoto,

J.: Toxicology of microcystins with reference to cases of human intoxication and epiemiological investigation of exposures to cyanobacteria and cyanotoxins. Archives of Toxicology (2016)

[14] Babica, P., Blaha, L., Marsalek, B.: Exploring the natural role of microcystins - a review of effects on photoautotrophic organisms. Journal of Phycology **42**(1), 9–20 (2006)

[15] Gan, N.Q., Xiao, Y., Zhu, L., Wu, Z.X., Liu, J., Hu, C.L., Song, L.R.: The role of microcystins in maintaining colonies of bloom-forming Microcystis spp. Environmental Microbiology **14**(3), 730–742 (2012)

[16] Boegehold, A.G., Burtner, A.M., Camilleri, A.V., Carter, G., DenUyl, P., Fanslow, D., Semenyuk, D., Godwin, C., Gossiaux, D., Johengen, T., Kelchner, H., Kitchens, C., Mason, L.A., McCabe, K., Palladino, D., Stuart, D., Vanderploeg, H., Errera, R.M.: Routine monitoring of western lake erie to track water quality changes associated with cyanobacteria harmful algal blooms. Earth System Science Data **15**(8), 3853–3868 (2023)

[17] Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.S.: A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics **2**(4), 1360–1383 (2008)

[18] Godage, N.H., Qian, S.S., Cudjoe, E., Gionfriddo, E.: Enhance quantitative analysis of Xenobiotics in blood plasma through cross-matrix calibration and Bayesian hierarchical modeling. ACS Measurement Science Au **4**(1), 127–135 (2024)

[19] Hunter, W.G., Lamboy, W.F.: A Bayesian analysis of the linear calibration problem. Technometrics **23**(4), 323–328 (1981)

[20] Stan Development Team: Stan Modeling Language User's Guide and Reference Manual, Version 2.30. (2022). http://mc-stan.org/manual.html

[21] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022). R Foundation for Statistical Computing. https://www.R-project.org/

[22] Stan Development Team: RStan: the R interface to Stan, Version 2.21.5 (2022). http://mc-stan.org/rstan.html

[23] Ott, W.R.: Environmental Statistics and Data Analysis. Lewis Publishers, Boca Raton, Florida (1995)

[24] Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA (1973)

[25] Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley, Chichester, UK (1994)

[26] Robbins, H.: An empirical Bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability vol. 1, pp. 157–163. The University of California Press, Berkeley, California (1956)

[27] Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B.: Shiny: Web Application Framework for R. (2023). https://shiny.posit.co/, https://github.com/rstudio/shiny