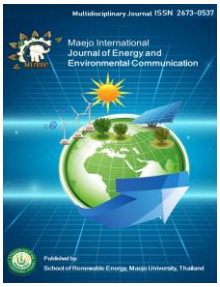
Maejo International Journal of Energy and Environmental Communication

Journal homepage: https://ph02.tci-thaijo.org/index.php/MIJEEC



ARTICLE

Application of hurdle Poisson model to predict the abundance of toxic cyanobacteria *Microcystis* in reservoirs

Truc-Ly Le-Huynh[1], Tomoaki Itayama[1]*, Kaito Mitsunaga[1], Misigo W. S. Angalika[1], Seiji Suzuki[1]

[1]Graduate School of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan.

ARTICLEINFO

Article history:
Received 24 November 2022
Received in revised form
13 December 2022
Accepted 15 December 2022

Keywords:
Toxic microcystis
Bayesian hurdle poisson model
mcyB gene
Air temperature
Trophic state index

ABSTRACT

The blooming of toxic cyanobacteria *Microcystis* in eutrophicated reservoirs causes serious difficulties for water supply worldwide. For the appropriate management of such reservoirs, a prediction model of toxic cyanobacteria *Microcystis* can be a useful tool. Therefore, this study aims to develop a Bayesian hurdle Poisson model for statistical prediction of toxic *Microcystis* from only two predictors, air temperature and trophic state index (TSI) calculated from chlorophyll-a. The gene copy number of the *mcyB* gene was used as a surrogate of toxic *Microcystis* cell density. The data on *mcyB* gene and chlorophyll-a were collected from 20 reservoirs in Nagasaki Prefecture (Japan). The daily average air temperature was downloaded from the local meteorological stations and a mean for 30 days before sampling date was calculated. The results showed that higher temperature and larger TSI accelerate the growth of toxic *Microcystis*. Furthermore, this model successfully predicted *mcyB* gene copy number as a surrogate of toxic *Microcystis* cell density for different conditions of air temperature and TSI with sufficient accuracy. Therefore, the proposed model has the potential to be a useful prediction tool for toxic cyanobacteria *Microcystis* in the integrated management of reservoirs.

1. Introduction

Cyanobacterial blooms have posed considerable challenges for water treatment and resource management worldwide (Chorus & Welker, 2021; Huisman et al., 2018). In recent decades, the blooms of cyanobacteria have been dramatically accelerated by eutrophication and climate change (Huisman et al., 2018; Paerl & Paul, 2012; Wagner & Adrian, 2009; Wiedner et al., 2007). In freshwater, cyanobacteria *Microcystis*, which can produce cyanotoxin microcystins (MCs), is one of the best-studied genera (Huisman et al., 2005). Prolonged exposure to MCs can cause liver damage, tumor promotion, and many chronic diseases (Chorus & Welker, 2021; Rastogi et al., 2014). Therefore, the development of predictive methods for the occurrence of toxic cyanobacteria, especially the toxic *Microcystis*, is valuable in reservoir water quality management to supply safe water.

Toxic *Microcystis* abundance in natural lakes and reservoirs has been determined by real-time PCR (polymerase chain reaction) using *mcy* gene (Kurmayer & Kutzenberger, 2003; Rantala-Ylinen et al., 2017). The real-time PCR method provides gene-specific, highly sensitive, and relatively fast results with objective quantification (Ibelings et al., 2014; Pacheco et al., 2016). As is well known, non-toxic *Microcystis* colonies, which do not carry microcystin synthesis genes (*mcy* genes), are frequently found together with toxic *Microcystis* colonies in many water bodies (Davis et al., 2009). However, it is impossible for microscopic observation to identify a toxic colony. Thus, PCR and real-time PCR based on a *mcy* gene detection have been applied (Kurmayer & Kutzenberger, 2003; Rantala-Ylinen et al., 2017).

* Corresponding author.
E-mail *address: itayama@nagasaki-u.ac.jp*

The statistical model is a common method for predicting the abundance of toxic cyanobacteria. Unlike the prediction method based on dynamical model, statistical model does not require a lot of continuous water quality data. Given the value of the ecosystem services of reservoirs against management cost, low-cost prediction methods would be applicable. Therefore, a statistical model that can predict toxic cyanobacteria with fewer predictors and less data would be highly beneficial. We considered the model based on the following criteria: i) predictors (explanatory variables) - water quality parameters that can be easily measured and accessed in the reservoir management; ii) variables that have been well confirmed to affect freshwater cyanobacterial blooms from previous studies. In addition, the survey results on toxic *Microcystis* can show a large number of detected as well as non-detected *mcy* gene in many reservoirs. Such a data set is called zero-inflated data, which has excessive numbers of zero counts and extreme positive values (Franks, 2018; Martin et al., 2005; Zuur et al., 2009). Such zero-inflated data can bias the analysis and give inaccurate results (Martin et al., 2005; Zuur et al., 2009). Therefore, statistical prediction models should directly take into account the zero-inflated data to properly estimate the model parameters. Cha et al. (2014) proposed the Bayesian hurdle Poisson model as a potential solution for zero-inflated cyanobacterial data. Bayesian analysis has been recommended as a promising approach in small-sample contexts (Kruschke, 2015; McNeish, 2016; Schoot & Miočević, 2020). In addition, it is easy to update the Bayesian statistical model by using the estimation results from past data as prior distributions and adding new data to estimate new posterior distributions. However, the study of Cha et al. (2014) could not quantify only toxic *Microcystis* cells, because Cha et al. (2014) applied morphological identification and enumeration of the cell under the optical microscope.

To predict the occurrence probability and abundance of toxic *Microcystis* from only two predictors (trophic state index (TSI) and air temperature), we studied a prediction method using the Bayesian hurdle Poisson model based on survey data of *mcyB* gene from 20 reservoirs in Nagasaki Prefecture, Japan. Statistical prediction models should directly take into account the zero-inflated data to properly estimate the model parameters. Cha et al. (2014) proposed the Bayesian hurdle Poisson model as a potential solution for zero-inflated cyanobacterial data. Bayesian analysis has been recommended as a promising approach in small-sample contexts (Kruschke, 2015; McNeish, 2016; Schoot & Miočević, 2020). In addition, it is easy to update the Bayesian statistical model by using the estimation results from past data as prior distributions and adding new data to estimate new posterior distributions. However, the study of Cha et al. (2014) could not quantify only toxic *Microcystis* cells, because Cha et al. (2014) applied morphological identification and enumeration of the cell under the optical microscope.

To predict the occurrence probability and abundance of toxic *Microcystis* from only two predictors (trophic state index (TSI) and air temperature), we studied a prediction method using the Bayesian hurdle Poisson model based on survey data of *mcyB* gene from 20 reservoirs in Nagasaki Prefecture, Japan.

2. Material and Methods

2.1 Data acquisition

2.1.1 Sampling sites

Water samples were collected from 20 reservoirs across Nagasaki Prefecture (Fig. 1) before and during the cyanobacterial blooming season (14/05/2017 – 13/05/2018, n = 42 observations). The surface water was carefully collected, but the blooming sample was avoided because a sample containing extremely high density of cyanobacteria is not representative.

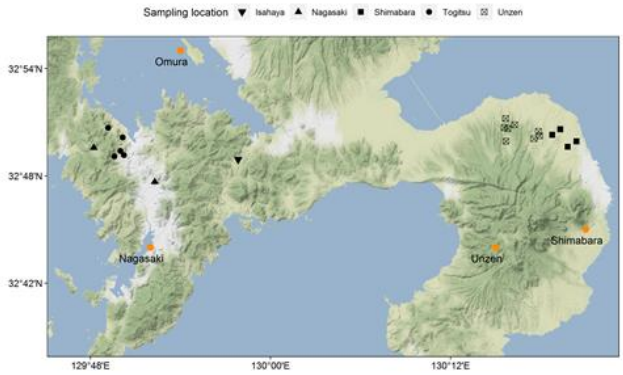


Figure 1 Sampling locations within Nagasaki Prefecture, Japan.

2.1.2 Trophic state index (TSI) from chlorophyll-a

In this study, TSI based on chlorophyll-a was used. Phytoplankton was collected by glass fiber filters of pore diameter less than 1µm (Advantec GA-55, Toyo Roshi Kaisha Ltd., Japan). Chlorophyll-a on the filter was measured using the hot methanol extraction and photometric method (Shimadzu UV-1800, Shimadzu Corporation, Kyoto, Japan) described by Saijo (1975). TSI value was calculated from chlorophyll-a concentration using the equation described by Carlson (1977).

$$TSI = 10\left(6 - \frac{2.04 - 0.68\ln(Chla)}{\ln(2)}\right) \qquad (1)$$

where $Chla$ is the chlorophyll-a concentration (µg/L) and ln ( ) is the natural logarithm function.

2.1.3 Air temperature

Data on daily temperature (°C) was available from local meteorological monitoring stations. The mean temperature for 30 days before the sampling date was calculated. Then the atmospheric temperature gradient (-0.6°C/+100 m) was used to determine the ambient temperature at each sampling site.

2.1.4 DNA extraction and mcyB gene quantification

The abundance of toxic *Microcystis* was quantified by *mcyB* gene using real-time PCR (qPCR) method. The genomic DNA in the reservoir samples of 35 mL was extracted by NucleoSpin kit (MACHEREY-NAGEL GmbH & Co. KG, Germany) following the manufacturer's manual. The *mcyB* gene was quantified by real-time PCR (qPCR) protocol as proposed by Sabart et al. (2010). The qPCR reaction mixture was performed with 10 µL of 2X Quantitec probe PCR kit mix (Qiagen), 900 nM *mcyB* primers, 250 nM *mcyB* TaqMan probe, and 1 µL of DNA template or external standard (*Microcystis* NIES843). Then sterile Milli-Q® water was added up to 20 µL. Thermal Cycler Dice Real-Time System (Takara Co., Japan) was used for the amplification with the following condition: initial preheating at 95°C for 15 min, followed by 50 cycles of 95°C

for 30 sec, 60°C for 1 min, and 72°C for 30 sec. No-template controls were included in each run. Since *mcyB* is a single-copy gene (Pacheco et al., 2016), this study assumed that *mcyB* gene copy number (copies/mL) was equivalent to toxic *Microcystis* cell density (cells/mL).

2.2 Statistical prediction model

In this study, the hurdle Poisson model analyzed toxic *Microcystis* data on absence (i.e., zero count) and presence (i.e., positive value) level by the binomial distribution. Then, the presence data were analyzed with the Poisson distribution. The model can be defined as:

$$P(y|\theta,\lambda) = \begin{cases} 1-\theta & : y = 0 \\ \theta \dfrac{Poisson\,(y|\lambda)}{1 - PoissonCDF(0|\lambda)} & : y > 0 \end{cases} \quad (2a)$$

$$ln(\lambda) = \beta_0 + \beta_1 \cdot \text{Temperature} + \beta_1 \cdot \text{TSI} + \nu \quad (2b)$$

where y is the gene copy number of mcyB (copies/mL); θ is the probability of mcyB presence in a sample; Poisson (|) and PoissonCDF(|) are Poisson distribution and the cumulative distribution function for Poisson distribution, respectively; λ is the mean parameter of the Poisson distribution; β_0 is the intercept, β_1 and β_2 are regression coefficients; ν is random intercept for sampling site; Temperature and TSI are converted from the air temperature data and the trophic state index in equation (1) using min-max normalization (minimum = 0, maximum = 1).

The hurdle Poisson model was analyzed within a Bayesian framework in R programming language (R Core Team, 2021) through package brms (Bürkner, 2017), which uses Markov chain Monte Carlo (MCMC) algorithm. The model was run for a total of 105 iterations with independent 4 chains of MCMC. The first 5×104 iterations were discarded as burn-in. The remaining were kept 1 in 10 samples (thinning) to reduce autocorrelation.

3. Results and discussion

3.1 Changes in trophic state index, air temperature, and toxic Microcystis (mcyB gene) abundance

The mean air temperature displayed a seasonal pattern and ranged from 11.4°C to 29.2°C as shown in Figure 2A. In Japan, the temperature is generally low in the early spring (March), then gradually increases in April – May and reaches the peak in the summer (July – August). The mean temperature in April (May) included March (April) data, which was lower than the temperature at the time of sampling. Considering the ecological succession of phytoplankton in reservoirs, it is clear that temperature at sampling time did not act as a predictor. Therefore, the mean was calculated from each daily averaged temperature for 30 days before the sampling date. The number of days to average is also a parameter for best estimation. However, it is necessary to compare the prediction performance of each model while changing the number of days, which requires large computational resources. Therefore, we decided to consider it in our next study.

TSI values, which were calculated from the chlorophyll-a concentration of 20 surveyed reservoirs (see equation (1)), ranged from 36.46 to 88.64. According to the classification by Carlson & Simpson (1996), the eutrophic reservoirs in this study accounted for 55% of the total, followed by hypertrophic reservoirs at 24%, and then mesotrophic reservoirs at 19%. Only one reservoir was classified as oligotrophic. More than three-quarters of surveyed reservoirs were in eutrophic and hypertrophic states, which are favorable conditions for the growth of cyanobacteria.

This study focused on the development of a prediction model of toxic *Microcystis* quantified by *mcyB* gene. The copy number of *mcyB* gene varied from 0 to nearly 6000 copies/mL (see Fig. 3). The proportion of zero counts in *mcyB* gene data was 50%, thus, the *mcyB* gene indicated a clear feature of zero-inflated data.

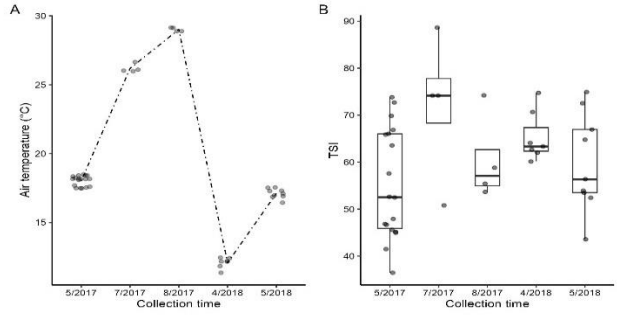


Figure 2 Variation in mean air temperature (A) and TSI (B).

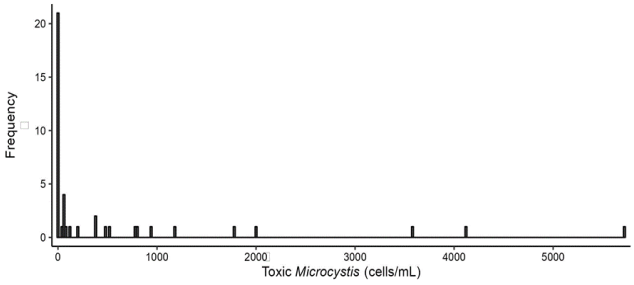


Figure 3 Distribution of mcyB gene copy number as toxic Microcystis cell density.

3.2 Estimated parameters of Bayesian hurdle poisson models

The MCMC parameter estimation of the Bayesian hurdle Poisson model was examined by trace plots, density plots, and German-Rubin statistics (Gelman & Rubin, 1992), though not shown here. As can be seen from the posterior distributions of the model in Table 1, the abundance of toxic *Microcystis* (*mcyB* gene) was positively associated with both air temperature and TSI. The estimated 95% credible interval (CI) of air temperature is [1.123, 1.235] and the 95% CI of TSI is [6.061, 6.423], and both intervals do not contain zero. If the 95% CI of the posterior distribution of a parameter contains zero, Bayesian statistics determines that the parameter is not significant (Kruschke, 2015). Therefore, the results imply that higher temperature and larger TSI positively affected the growth of toxic *Microcystis*. Furthermore, since the TSI and temperature data are standardized, the estimated parameter magnitudes can be compared with each other. The median of the posterior distribution of the air temperature is 1.180 and that of TSI is 6.236. This result clearly suggests that TSI was about 6 times higher than air temperature in terms of its effect on toxic *Microcystis* growth. On the other hand, the estimated median of the hurdle probability is 0.500, but the estimated 95% CI is [0.353, 0.644], which is relatively higher than the other estimation errors. The hurdle probability can be interpreted as the occurrence

probability of toxic *Microcystis* in a reservoir. Whether *Microcystis* is toxic or non-toxic depends on the presence or absence of the MCs synthesis gene (Huisman et al., 2005). This varies from strain to strain of *Microcystis* and is genetically determined (Huisman et al., 2005). Therefore, the presence of toxic *Microcystis* strains in a reservoir can in principle be considered as a probabilistic event regardless of the water quality of the reservoir. The probability may be expected to differ from region to region.

These results indicate that the hurdle Poisson model is a good model for predicting cyanobacteria in water. Cha et al. (2014) already elucidated that Bayesian hurdle Poisson model can be a practical tool for handling zero-inflated cyanobacterial data, but they could not distinguish between toxic and non-toxic cyanobacteria due to the disadvantage of microscopic observation. In contrast, our study is the first successful application of the Bayesian hurdle Poisson model in predicting toxic *Microcystis* using *mcyB* gene copy number.

Table 1 Median, standard error, and 95% credible intervals in posterior distributions of the regression coefficients in the model.

	Estimate (median)	Estimate Error	95% Credible Intervals	Rhat
Intercept (β_0)	1.600	0.417	[0.698, 2.477]	1.00
Temperature (β_1)	1.180	0.028	[1.123, 1.235]	1.00
TSI (β_2)	6.236	0.096	[6.051, 6.423]	1.00
Hurdle probability (θ)	0.500	0.076	[0.353, 0.644]	1.00

Note:

Rhat is the potential scale reduction factor on split chains; Rhat = 1 at convergence.

3.3 Predicting toxic microcystis abundance

The purpose of this study is to predict the proliferation of toxic *Microcystis* evaluated by *mcyB* gene. The probabilistic prediction of *mcyB* gene copy number can be computed by using the estimated posterior distribution of parameters in the hurdle Poisson model shown in Table 1. Figure 4 demonstrated the effects of temperature and TSI on the proliferation of toxic *Microcystis* evaluated by *mcyB* gene copy number. Each shaded area in Fig. 4 shows 80% prediction interval. Because high temperature is the optimal condition for the proliferation of *Microcystis*, especially toxic *Microcystis* (Chorus & Welker, 2021), we selected 20°C, 25°C, and 30°C as displayed temperatures to represent the possible worst blooming scenarios. Besides, we chose two thresholds from the WHO alert level framework for drinking water to illustrate the application of our proposed model in reservoir management. The threshold of Vigilance Level is 200 cells/mL, and the threshold of Alert Level 1 is 2000 cells/mL (Bartram & Chorus, 1999). At 20°C, toxic *Microcystis* might exceed Vigilance Level and Alert Level 1 at the TSI of 67 and 87, respectively. At 25°C, those TSI values are 64 and 84; and at 30°C, those TSI values are 61 and 81. In oligotrophic and mesotrophic reservoirs (TSI < 50), toxic *Microcystis* densities are certainly below Vigilance Level. Toxic *Microcystis* exceeded Alert Level 1 only in hypertrophic reservoirs with TSI greater than 80, and it was expected that nutrient load reduction measures would be required. Also, in eutrophic reservoirs with a TSI greater than 60, water quality monitoring will need to be enhanced for the early warning of toxic *Microcystis*.

In this study, the Bayesian hurdle Poisson model successfully predicted the copy number of *mcyB* gene (as a surrogate for toxic *Microcystis*) with sufficient accuracy, even though only TSI and temperature were used as predictors. Air temperature data can be easily downloaded from local meteorological stations. The measurement of chlorophyll-a is relatively fast, cheap, and easy (Ibelings et al., 2014), suggesting that TSI calculated from chlorophyll-a is appropriate for the predictor. Of course, it is difficult to use the parameters estimated from reservoirs in Nagasaki Prefecture for reservoirs far from this area. If we want to establish a prediction model for regional reservoirs in Thailand or other Southeast Asian countries, we need to collect data from reservoirs in or near the target area. However, the results of this study allow us to conclude that the Bayesian hurdle Poisson model using TSI and air temperature as predictors is useful for predicting toxic cyanobacteria in the practical management of reservoirs.

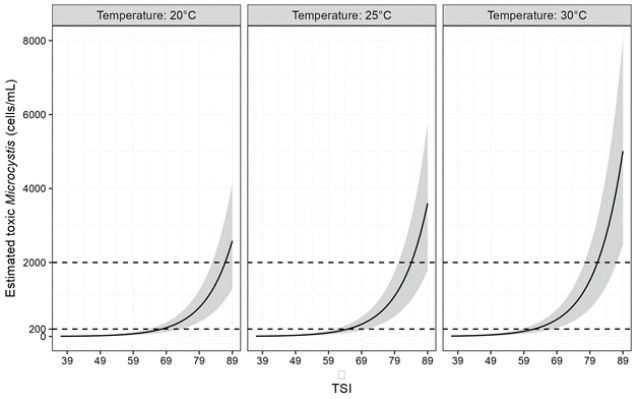


Figure 4 Estimated toxic Microcystis abundance (cells/mL) (with 80% prediction interval – shaded area) as a function of TSI at different temperature (20°C, 25°C, 30°C). The dashed lines are WHO alert thresholds for drinking water, Vigilance Level (200 cells/mL) and Alert Level 1 (2000 cells/mL).

4. Conclusion

Bayesian hurdle Poisson model successfully handled zero-inflated data of *mcyB* gene, which is used as a surrogate of toxic *Microcystis*, surveyed from 20 reservoirs in Nagasaki prefecture, Japan. MCMC calculation for this model estimated the posterior distribution of TSI and air temperature, which were predictors in the model. The obtained results elucidated that higher temperature and larger TSI accelerate the growth of toxic *Microcystis*. Furthermore, it is suggested that TSI was about 6 times higher than the air temperature in terms of its effect on toxic *Microcystis* growth. This model successfully predicted *mcyB* gene copy number as a surrogate of toxic *Microcystis* cell density for different conditions of air temperature and TSI with sufficient accuracy. In oligotrophic and mesotrophic reservoirs, toxic *Microcystis* densities are certainly below the Vigilance Level. Also, in eutrophic reservoirs with a TSI greater than 60, intensive water quality monitoring is required for the early warning of toxic *Microcystis*.

Reference

Bartram, J., & Chorus, I. (Eds.). (1999). Toxic Cyanobacteria in Water: A Guide to their Public Health Consequences, Monitoring and Management (1st ed.). CRC Press.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1–28.

Carlson, R. E. (1977). A trophic state index for lakes1. Limnology and Oceanography, 22(2), 361–369.

Carlson, R., & Simpson, J. (1996). A Coordinator's Guide to Volunteer Lake Monitoring Methods. https://www.nalms.org/secchidipin/monitoring-methods/trophic-state-equations/

Cha, Y., Park, S. S., Kim, K., Byeon, M., & Stow, C. A. (2014). Probabilistic prediction of cyanobacteria abundance in a Korean reservoir using a Bayesian Poisson model. Water Resources Research, 50(3), 2518–2532.

Chorus, I., & Welker, M. (Eds.). (2021). Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management (2nd ed.). CRC Press.

Davis, T. W., Berry, D. L., Boyer, G. L., & Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae, 8(5), 715–725.

Franks, P. J. S. (2018). Recent Advances in Modelling of Harmful Algal Blooms. In P. M. Glibert, E. Berdalet, M. A. Burford, G. C. Pitcher, & M. Zhou (Eds.), Global ecology and oceanography of harmful algal blooms (pp. 359–377). Springer, Cham.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7(4), 457–472.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. (2018). Cyanobacterial blooms. Nature Reviews Microbiology, 16(8), 471–483.

Huisman, J., Matthijs, H. C., & Visser, P. M. (Eds.). (2005). Harmful Cyanobacteria (1st ed., Vol. 3). Springer-Verlag.

Ibelings, B. W., Backer, L. C., Kardinaal, W. E. A., & Chorus, I. (2014). Current approaches to cyanotoxin risk assessment and risk management around the globe. Harmful Algae, 40, 63–74.

Kruschke, J. K. (2015). Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan (2nd ed.). Elsevier. https://doi.org/10.1016/B978-0-12-405888-0.09999-2

Kurmayer, R., & Kutzenberger, T. (2003). Application of Real-Time PCR for Quantification of Microcystin Genotypes in a Population of the Toxic Cyanobacterium Microcystis sp. Applied and Environmental Microbiology, 69(11), 6723–6730.

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J., & Possingham, H. P. (2005). Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. Ecology Letters, 8(11), 1235–1246.

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. Structural Equation Modeling: A Multidisciplinary Journal, 23(5), 750–773.

Pacheco, A. B. F., Guedes, I. A., & Azevedo, S. M. (2016). Is qPCR a reliable indicator of cyanotoxin risk in freshwater? Toxins, 8(6).

Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. Water Research, 46(5), 1349–1363.

R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/

Rantala-Ylinen, A., Savela, H., Sivonen, K., & Kurmayer, R. (2017). Quantitative PCR. In R. Kurmayer, K. Sivonen, A. Wilmotte, & N. Salmaso (Eds.), Molecular tools for the detection and quantification of toxigenic cyanobacteria (pp. 205–239). John Wiley & Sons, Ltd.

Rastogi, R. P., Sinha, R. P., & Incharoensakdi, A. (2014). The cyanotoxin-microcystins: Current overview. Reviews in Environmental Science and Biotechnology, 13(2), 215–249.

Sabart, M., Pobel, D., Briand, E., Combourieu, B., Salençon, M. J., Humbert, J. F., & Latour, D. (2010). Spatiotemporal Variations in Microcystin Concentrations and in the Proportions of Microcystin-Producing Cells in Several Microcystis aeruginosa Populations. Applied and Environmental Microbiology, 76(14), 4750–4759.

Saijo, Y. (1975). A method for determination of chlorophyll. Japanese Journal of Limnology, 36(3), 103–109.

Schoot, R. van de, & Miočević, M. (Eds.). (2020). Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners (1st ed.). Routledge.

Wagner, C., & Adrian, R. (2009). Cyanobacteria dominance: Quantifying the effects of climate change. Limnology and Oceanography, 54(6, part 2), 2460–2468.

Wiedner, C., Rücker, J., Brüggemann, R., & Nixdorf, B. (2007). Climate change affects timing and size of populations of an invasive cyanobacterium in temperate regions. Oecologia, 152(3), 473–484.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. Springer New York.