Water Research 176 (2020) 115710

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Predicting cyanobacterial biovolume from water temperature and conductivity using a Bayesian compound Poisson-Gamma model

Signe Haakonsson [a, b, *], Marco A. Rodríguez [c], Carmela Carballo [a], María del Carmen Pérez [a, b], Rafael Arocena [a], Sylvia Bonilla [a, b]

[a] Sección Limnología, Instituto de Ecología y Ciencias Ambientales, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400, Montevideo, Uruguay
[b] Physiology and Ecology Phytoplankton Group, CSIC 1176, Uruguay
[c] Département des Sciences de l'environnement, Université du Québec à Trois-Rivières, 3351 boulevard des Forges, Trois-Rivières, Québec, G9A 5H7, Canada

ARTICLE INFO

Article history:
Received 12 September 2019
Received in revised form
6 March 2020
Accepted 12 March 2020
Available online 19 March 2020

Keywords:
Alert levels
Bayesian modeling
Brackish ecosystems
HAB management
Risk thresholds
Salinity

ABSTRACT

Eutrophication and climate change scenarios engender the need to develop good predictive models for harmful cyanobacterial blooms (CyanoHABs). Nevertheless, modeling cyanobacterial biomass is a challenging task due to strongly skewed distributions that include many absences as well as extreme values (dense blooms). Most modeling approaches alter the natural distribution of the data by splitting them into zeros (absences) and positive values, assuming that different processes underlie these two components. Our objectives were (1) to develop a probabilistic model relating cyanobacterial biovolume to environmental variables in the Río de la Plata Estuary (35°S, 56°W, n = 205 observations) considering all biovolume values (zeros and positive biomass) as part of the same process; and (2) to use the model to predict cyanobacterial biovolume under different risk level scenarios using water temperature and conductivity as explanatory variables. We developed a compound Poisson-Gamma (CPG) regression model, an approach that has not previously been used for modeling phytoplankton biovolume, within a Bayesian hierarchical framework. Posterior predictive checks showed that the fitted model had a good overall fit to the observed cyanobacterial biovolume and to more specific features of the data, such as the proportion of samples crossing three threshold risk levels (0.2, 1 and 2 $mm^3\ L^{-1}$) at different water temperatures and conductivities. The CPG model highlights the strong control of cyanobacterial biovolume by nonlinear and interactive effects of water temperature and conductivity. The highest probability of crossing the three biovolume levels occurred at 22.2 °C and at the lowest observed conductivity (~0.1 mS cm^{-1}). Cross-validation of the fitted model using out-of-sample observations (n = 72) showed the model's potential to be used *in situ*, as it enabled prediction of cyanobacterial biomass based on two readily measured variables (temperature and conductivity), making it an interesting tool for early alert systems and management strategies. Furthermore, this novel application demonstrates the potential of the Bayesian CPG approach for predicting cyanobacterial dynamics in response to environmental change.

1. Introduction

Harmful cyanobacterial blooms (CyanoHABs) have been registered with increasing frequency in many aquatic habitats worldwide during recent decades (Carey et al., 2012; Huisman et al., 2018). Eutrophication and climate change are the major factors underlying the escalation of the extent and duration of CyanoHABs (Bonilla and Pick, 2017; O'Neil et al., 2012). The toxins that may be contained in these blooms threaten supplies of drinking water, endanger human activities and have negative effects on fish populations that can imperil fisheries (Dodds et al., 2009; Steffensen, 2008). Consequently, the prediction of cyanobacterial occurrence and biomass has become critical both to developing strategies preventing CyanoHAB formation and to reducing the impact of blooms already underway (Glibert et al., 2010). However, modeling cyanobacteria is a challenging task due to the dynamic nature of their biomass, which changes rapidly in space and time (Oliver

* Corresponding author. Sección Limnología, Instituto de Ecología y Ciencias Ambientales, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400, Montevideo, Uruguay.
E-mail address: shaakonsson@fcien.edu.uy (S. Haakonsson).

https://doi.org/10.1016/j.watres.2020.115710
0043-1354/

et al., 2012).

The dominance of cyanobacteria in the phytoplankton of freshwater and brackish ecosystems is related to climate, hydrology and eutrophication. Favorable conditions that may trigger blooms include high water temperature (Beaulieu et al., 2013; Elliott, 2010; Haakonsson et al., 2017; Paerl and Huisman, 2008), low salinity (Engström-Öst et al., 2011), high nutrient concentrations (Downing et al., 2001; Gobler et al., 2016; Smith, 1986; Smith and Schindler, 2009), high light penetration (Davis and Koop, 2006) and water column stability (Carey et al., 2012; Wagner and Adrian, 2009). However, the variables that best explain occurrence are not always those that best predict it (Shmueli, 2010). In eutrophic ecosystems, nutrients (phosphorus, P, and nitrogen, N) are major determinants of CyanoHABs (Lancelot and Muylaert, 2011; Lehman et al., 2010), but other variables, such as water temperature, salinity and hydrological features, can be more effective as predictors of bloom dynamics (Olli et al., 2015; Robson and Hamilton, 2004; Taş et al., 2006). Measuring nutrient concentrations is time-consuming, which limits their utility as predictors for early warning systems. Ideally, such systems would include environmental features that can be measured rapidly.

Prediction of planktonic cyanobacterial biomass and abundance has traditionally relied on a variety of quantitative methods, such as linear and nonlinear regression (Downing et al., 2001; Kosten et al., 2012; Persaud et al., 2015), additive models (Beaulieu et al., 2013; Taranu et al., 2012), regression trees (Ghaffar and Stevenson, 2016), simulation models (Elliott, 2010) and artificial neural networks (Ahn et al., 2011; Yang et al., 2016). More recent approaches include fully probabilistic models such as Bayesian networks (Johnson et al., 2010; Rigosi et al., 2015) and Bayesian regression analysis (Cha et al., 2014; Hamilton et al., 2009; Obenour et al., 2014). Bayesian approaches have high predictive power and allow for realistic quantification of uncertainty and separation of measurement and process errors (Franks, 2018; Hamilton et al., 2009), which are key features for prediction. A Bayesian probabilistic approach is useful for risk management applications such as, for example, predicting cyanobacterial biovolume exceedance of different risk exposure levels (Cha et al., 2014).

Cell abundance is the most used response variable in predictive CyanoHAB models (e.g., Cha et al., 2017; Ghaffar and Stevenson, 2016) whereas biovolume, a more accurate indicator of biomass, has been less frequently used (Dolman et al., 2012; Kosten et al., 2012). The reasons behind this trend likely include the lesser availability of biovolume data (as a result of the great effort involved in measuring the volume of each taxon). Additionally, biovolume is a continuous variable with a skewed distribution that includes many zeros as well as extreme values, all of which pose challenges for modeling. Species data often contain a large proportion of zeros (Fletcher et al., 2005). To deal with this problem there are two common approaches: adding a small arbitrary value to eliminate the zeros (e.g.: $x + 0.1$), or using two-part "hurdle" models (one for the zeros, the other for the positive values); the latter approach requires an artificial split in the distribution of the data. Hurdle models have been used to predict microcystin concentrations in lakes and reservoirs (Taranu et al., 2017). An alternative approach, based on the compound Poisson-Gamma (CPG) distribution, allows for joint modeling of zeros and strictly positive values of a continuous variable, and for extreme values, such as those arising from blooms (Lecomte et al., 2013a; Pennington, 1996). The CPG distribution has been used by ecologists to model biomass of fishery catches (Foster and Bravington, 2013; Lecomte et al., 2013a) and benthic macroinvertebrates (Lecomte et al., 2013b). Despite its suitability for dealing with ecological data, to our knowledge, the CPG approach has not yet been used to model phytoplankton biovolume.

The selection of a suitable indicator is critical to a successful risk prediction. Most regulatory and advisory guidelines for drinking and recreational waters define levels or risk exposure to cyanobacteria based on some kind of concentration indicator (Chorus and Bartram, 1999). The most commonly used indicators are chlorophyll *a* concentration and cyanobacterial cell abundance or biovolume (Chorus and Bartram, 1999). Among these three variables, biovolume is the most accurate for estimating population biomass, since cell size can vary more than one order of magnitude within species and more than two between species (Chorus and Cavalieri, 2000), and toxin concentration is better correlated with biomass than with cell counts (Ibelings et al., 2015). For example, the World Health Organization proposed two biovolume alert levels for drinking water (0.2 and 10 $mm^3\ L^{-1}$) (Chorus and Bartram, 1999), while many countries have implemented two or three guide levels (varying from <0.5 to >15 $mm^3\ L^{-1}$) for recreational waters, with associated consequent actions (Ibelings et al., 2015).

The Río de la Plata Estuary, one of the largest in the world (35000 km^2), is located in the subtropical region of southern South America (Fig. 1). The estuary provides multiple ecosystem services such as transportation, fisheries and tourism, and is the main drinking water supply for Buenos Aires (20 million inhabitants) (Giannuzzi et al., 2012; Nagy et al., 2002). The estuary has shown symptoms of eutrophication for decades (Nagy et al., 2002) and blooms of toxic cyanobacteria have been reported since 1999 (De León and Yunes, 2001). The blooms occur with high frequency especially in summer along both the south (Aguilera et al., 2017; Giannuzzi et al., 2009; Gómez, 2014; Sathicq et al., 2014) and the north shores (Bonilla et al., 2015; Brena et al., 2006; De León and Yunes, 2001; Kruk et al., 2015), including recently (2019) observed massive blooms (Kruk et al., 2019). Changes in physical variables such as temperature and salinity appear to be directly related to these blooms, and thus the development of predictive models based on these variables should lead to improved water monitoring programs. Our objectives were to develop a compound Poisson-Gamma regression model relating cyanobacterial biovolume to readily measured environmental variables, and to use the model within a full Bayesian framework to predict the probabilities of reaching three biovolume thresholds: 0.2 $mm^3\ L^{-1}$, 1 $mm^3\ L^{-1}$, and 2 $mm^3\ L^{-1}$. Because there is no current regulatory framework for recreational waters in Uruguay, we selected these thresholds based on globally accepted risk exposure levels (Chorus and Bartram, 1999; Ibelings et al., 2015) and observed cyanobacterial dynamics in the study region (Bonilla et al., 2015; Haakonsson et al., 2017). We examined cyanobacterial abundances over a five-year period at three sampling points ($n = 277$ observations) located on the shore and 500 m offshore in the Río del Plata Estuary.

2. Material and methods

2.1. Data acquisition

Phytoplankton and environmental variables were sampled at Punta del Tigre in the Río de la Plata Estuary, Uruguay (35°45′56.0″S, 56°33′27.7″W), on a bimonthly (spring and summer) and trimonthly (autumn and winter) basis, between October 2014 and July 2019 (2014–2018, $n = 205$ observations used to calibrate the model; 2018–2019, $n = 72$ observations used for cross-validation). The study area is shallow (<4 m depth), oligohaline and meso-eutrophic (total phosphorus: 118 μg L^{-1} and 81–144 μg L^{-1}; chlorophyll *a*: 5 μg L^{-1} and 2–10 μg L^{-1}, medians and 25%–75% quartiles, respectively, $n = 24$; Rafael Arocena, unpublished data). Samples were collected from three sites, one at the shore (sand beach, mean depth = 0.4 m, site 1) and two located

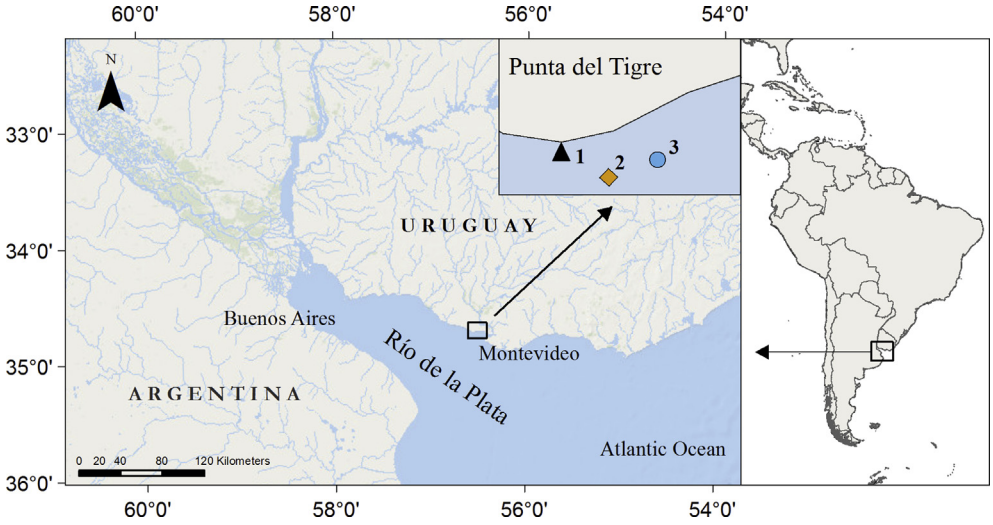

Fig. 1. Location of the three sampling sites (1–3) at Punta del Tigre on the north coast of the Río de la Plata Estuary, Uruguay.

500 m offshore (mean depth = 3.5 m and 3.8 m, sites 2 and 3, respectively) (Fig. 1). Depth, water temperature, conductivity (a proxy for salinity), dissolved oxygen, pH and light penetration were measured at each sampling site.

Phytoplankton samples were taken (in triplicate at most sites and dates; mean number of replicates = 2.71) from the subsurface at each site in 0.5 L bottles and fixed in Lugol's solution. Algae and cyanobacteria taxa were identified to the lowest possible taxonomical level and counted (cells mL^{-1}) in sedimentation chambers using an inverted microscope (Olympus CKX41) and a standard quantification method (Sournia, 1978; Utermöhl, 1958). Taxa volumes (μm^3) were calculated from geometric shape approximations (Hillebrand et al., 1999) using cell measurements (at least 10 cells per taxon, per sample) to obtain the biovolume of each taxon ($mm^3\ L^{-1}$).

2.2. Model development

The Compound Poisson-Gamma (CPG) variables are sums of random variables arising from a common Gamma distribution, in which the number of terms entering the sums has a Poisson distribution. In our application, the total cyanobacterial biovolume in a sample (Y) is viewed as a collection of M cyanobacterial aggregations (groups of organisms viewed as statistical "events", the number of which is sampled from a Poisson distribution), having biovolumes R which follow a Gamma distribution (Lecomte et al., 2013a). More formally, $Y = R_1 + R_2 + R_3 + \ldots + R_M$, where M has a Poisson distribution with mean λ, and the R are identically distributed random variables sampled from a Gamma distribution with shape α and rate β (Withers and Nadarajah, 2011). This implies that there may be samples with no cyanobacteria ($M = 0$). Regression models can be cast in this framework by modeling the Poisson mean as a function of covariates (Lecomte et al., 2013a; Smyth, 1996). In these models, the information in both the zero and positive observations can contribute to the estimation of all parts of the model (Smyth, 1996).

We developed a CPG regression model within a full Bayesian framework using cyanobacterial biovolume as the response variable. The model includes two environmental variables that strongly influence cyanobacterial biomass in coastal ecosystems: surface water temperature and conductivity, a proxy for salinity. Previous studies in coastal ecosystems worldwide (Olli et al., 2015; Robson and Hamilton, 2004; Taş et al., 2006) and in the Río de la Plata Estuary (Kruk et al., 2019; Martínez de la Escalera et al., 2017) have revealed the importance of these two variables for bloom prediction. Preliminary tests with models including pH, flow and depth as possible predictors indicated that these variables had no discernible effect on cyanobacterial biovolume prediction; these variables were thus omitted from further analyses.

Our model assumes that the biovolume observations Y arise from a CPG distribution characterized by three parameters: the Poisson mean (λ), and the Gamma shape (α) and rate (β):

$$Y_{ijk} \sim \mathrm{CPG}(\lambda_{ijk}, \alpha, \beta),$$

where the indices correspond to replicate i ($i = 1, ..., N_{jk}$), at site j ($j = 1, 2, 3$) and time k ($k = 1, \ldots, 26$). The number of replicates N_{jk} is indexed by site and date because samples were not collected in triplicate at all sites and dates (see 2.1 *Data acquisition*).

The Poisson mean, λ_{ijk}, was modeled on the log scale as:

$$\log(\lambda_{ijk}) = \phi_0 + \phi_1 X_{1ijk} + \phi_2 X_{2ijk} + \ldots + \phi_p X_{pijk} + \eta_{ij} + \nu_{ik} + \varepsilon_{ijk}$$

where ϕ_0 is an intercept and $\phi_1, \ldots, \phi_p$ are slope parameters for p covariates $X_{1ijk}, \ldots, X_{pijk}$; η_{ij} are spatial random effects associated with differences among sites; ν_{ik} are temporal random effects associated with differences among dates; and ε_{ijk} are observation-level errors. Our model included four covariate terms that captured additive and multiplicative effects of the explanatory variables temperature and conductivity: linear and quadratic terms for water temperature, a linear term for conductivity, and a term for the interaction between temperature and conductivity. Inclusion of a quadratic term for temperature allows for dome-shaped relationships between biovolume and temperature, including a thermal optimum at intermediate temperatures, whereas the temperature-conductivity interaction allows the effect of either variable to be affected by the level of the other. Temperature and conductivity were standardized to zero mean and unit variance to facilitate comparisons of effects and improve computational stability.

We used the following diffuse prior distributions for the model parameters:

Regression coefficients:

$$\phi_i \sim \mathrm{Normal}(0, 100)$$

Gamma shape and rate parameters:

$\alpha \sim \text{Gamma}(0.01, 0.01)$

$\beta \sim \text{Gamma}(0.01, 0.01)$

Random effects:

$\eta_{ij} \sim \text{Normal}(0, \sigma^2_\eta)$

$\nu_{ik} \sim \text{Normal}(0, \sigma^2_\nu)$

$\sigma^2_\eta \sim \text{Normal}_+(0, 10)$

$\sigma^2_\nu \sim \text{Normal}_+(0, 10)$

Observation error:

$\varepsilon_{ijk} \sim \text{Normal}\ (0, \sigma^2_\varepsilon)$

$\sigma^2_\varepsilon \sim \text{Normal}_+(0, 10)$

We used prior-posterior comparisons to gauge the information gained from the data relative to the vague priors. We used posterior predictive checks, which compare the observed data to the posterior predictive distribution (Gelman et al., 1996, 2014), to assess both the overall goodness-of-fit of the model and its ability to predict specific aspects of the data which are of particular interest for management.

To conduct these tests, we devised three data-dependent test quantities, calculated as the proportion of samples that exceeded each of three selected cyanobacterial biovolume thresholds (0.2 $mm^3\ L^{-1}$, 1 $mm^3\ L^{-1}$, and 2 $mm^3\ L^{-1}$). The first threshold corresponds to the "Alert Level 1" for drinking waters and the last to the "Guidance Level 1" for recreational waters of the WHO alert level framework (Chorus and Bartram, 1999). The intermediate threshold value (1 $mm^3\ L^{-1}$), used in different countries for recreational waters (Ibelings et al., 2015), corresponds to a biomass typically found prior to visible accumulations (i.e., 2 $mm^3\ L^{-1}$ and beyond) and is therefore useful for developing early warning systems. The three test quantities calculated from the observed data were compared to those generated by simulations from the posterior predictive distribution. We calculated variance components for the random terms in the model to quantify the relative contribution of spatial and temporal variation to fluctuations in biovolume. The fitted model was also used to obtain the probability of exceeding each of the selected biovolume thresholds at different combinations of temperature and conductivity, taken as the proportion of MCMC samples from the posterior distribution of biovolume that had values greater than the threshold.

All analyses were carried out in the R environment (R Core Team, 2018) using the rjags interface (Plummer, 2014) to JAGS (Plummer, 2003), which uses a MCMC algorithm to sample from the posterior distribution of the parameters. An initial adaptive phase was run for 1,000 MCMC iterations to optimize sampler efficiency by tuning the proposal distribution. We ran 5 chains for 2,000,000 iterations, discarded the first half (burn-in), and retained 1 in 1,000 samples (thinning) of the second half to obtain a total of 5,000 MCMC samples. Posterior distributions of individual parameters were summarized as medians and 95% credible intervals. Convergence of the chains was assessed using visual inspection of trace plots and the univariate and multivariate Gelman–Rubin diagnostics (potential scale reduction factors, PSRF) in the CODA package (Plummer et al., 2006).

In addition to the posterior predictive checks, we cross-validated the fitted model against out-of-sample observations (2018–2019; n = 72) which were not used to fit the model. The model fitted to the first set of observations (2014–2018) was used to predict cyanobacterial biovolume from temperature and conductivity in the second set of observations (2018–2019). Predictions for new observations were obtained after setting all random effects to zero (these are conditional responses for observations in a hypothetical sampling group; Skrondal, 2009).

3. Results

3.1. Dynamics of cyanobacteria and main predictive variables

The three sampling sites were similar in their limnological characteristics during the study period (Table 1, Fig. 2). Temperature ranged from 9.5 to 29.2 °C, showing a clear seasonal trend, whereas conductivity ranged from 0.1 to 8.7 mS cm^{-1} and showed no clear temporal trend (Fig. 2). Cyanobacterial biovolume ranged from 0 $mm^3\ L^{-1}$ (101 of 205 samples) to >60 $mm^3\ L^{-1}$ (Fig. 3). At low biovolume (<0.2 $mm^3\ L^{-1}$), the phytoplankton assemblage was mainly dominated by diatoms (such as *Aulacoseira granulata* and *Actinocyclus normanii*) followed by dinoflagellates, cryptophytes and chlorophytes. During the warm season, when the total biovolume was moderate to high (>1 $mm^3\ L^{-1}$), the assemblage was characterized by cyanobacteria (>60% of total phytoplankton biovolume). The most dominant species were the colonial *Microcystis aeruginosa* and *M. protocystis* and the filamentous *Dolichospermum uruguayensis* and *D. circinale*. A complete list of cyanobacterial taxa is presented in the Supplementary Material (Table S1).

3.2. Fit and parameter estimates of the CPG model

Neither the multivariate (1.07) nor the univariate Gelman-Rubin diagnostics (<1.04 for all parameters) indicated potential problems with convergence, i.e., chains seemed to mix adequately, as the Gelman-Rubin diagnostic was <1.1 (Gelman et al., 2014). The posterior predictive distribution of cyanobacterial biovolumes showed a good overall fit to the observed data (Fig. 4a, Fig. S1 Supplementary Material). Posterior predictive checks showed that model predictions were within a few percentage points of the observed values for the three alert level thresholds, although the model may be slightly underestimating the proportion of samples exceeding the lowest threshold (0.2 $mm^3\ L^{-1}$; Fig. 4b). Comparison of prior and posterior distributions (Fig. 5) showed that the data provided substantial information on the parameters of interest and that parameter estimates were not determined primarily by the prior distribution.

The parameter estimates associated with the covariates in our model, surface temperature (ϕ_1), surface temperature^2 (ϕ_2), conductivity (ϕ_3), and the interaction between surface temperature and conductivity (ϕ_4), showed a clear effect of all covariates on cyanobacterial biovolume (Fig. 5), and point to the presence of nonlinear responses to temperature and conductivity. The components of variance yielded by the model for biovolume indicated that temporal variation was substantially stronger than spatial variation (Supplementary Material, Table S2), matching the patterns observed earlier for the environmental predictors (Fig. 2).

The cross-validation showed that the model captured the main seasonal trends at the three study sites, although the wide credible intervals point to substantial variability in the predicted values (Fig. 6). With the exception of the mid-summer (Nov.–Dec.) samples, there appears to be no systematic over- or under-estimation of mean biovolume across sites. The coefficients of determination for observed and predicted values at the three sites were: $r^2 = 0.37$, for site 1, $r^2 = 0.34$ for site 2 and $r^2 = 0.92$ for site 3 (n = 8 observations

Table 1
Environmental characteristics (medians; 25%–75% quartiles) at the three sampling sites (sample sizes in parentheses).

Site	Depth (m)	pH	Dissolved oxygen (mg/L)	Turbidity (NTU)
1	0.4; 0.3–0.6 (26)	8.3; 8.0–8.5 (26)	9.2; 7.3–11.3 (26)	–
2	3.4; 3.2–3.7 (25)	7.9; 7.9–8.2 (25)	9.5; 8.7–10.2 (25)	27; 20.6–42 (25)
3	3.8; 3.6–4.1 (25)	7.9; 7.7–8.2 (25)	9.4; 8.5–9.7 (25)	25; 20.2–42 (25)

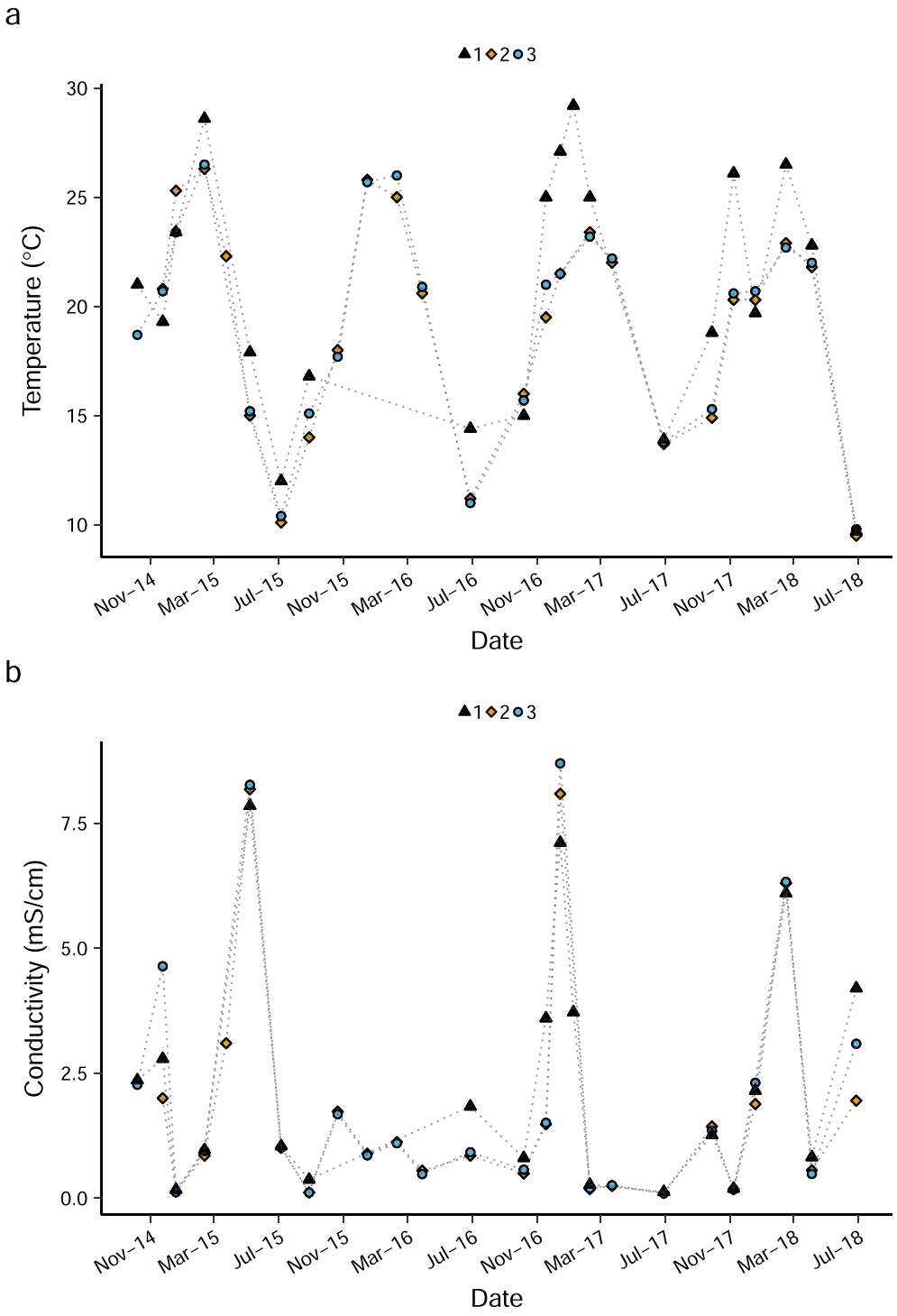

Fig. 2. Temporal and spatial variation in temperature (a) and conductivity (b) at the three sampling sites (site numbering as in Fig. 1).

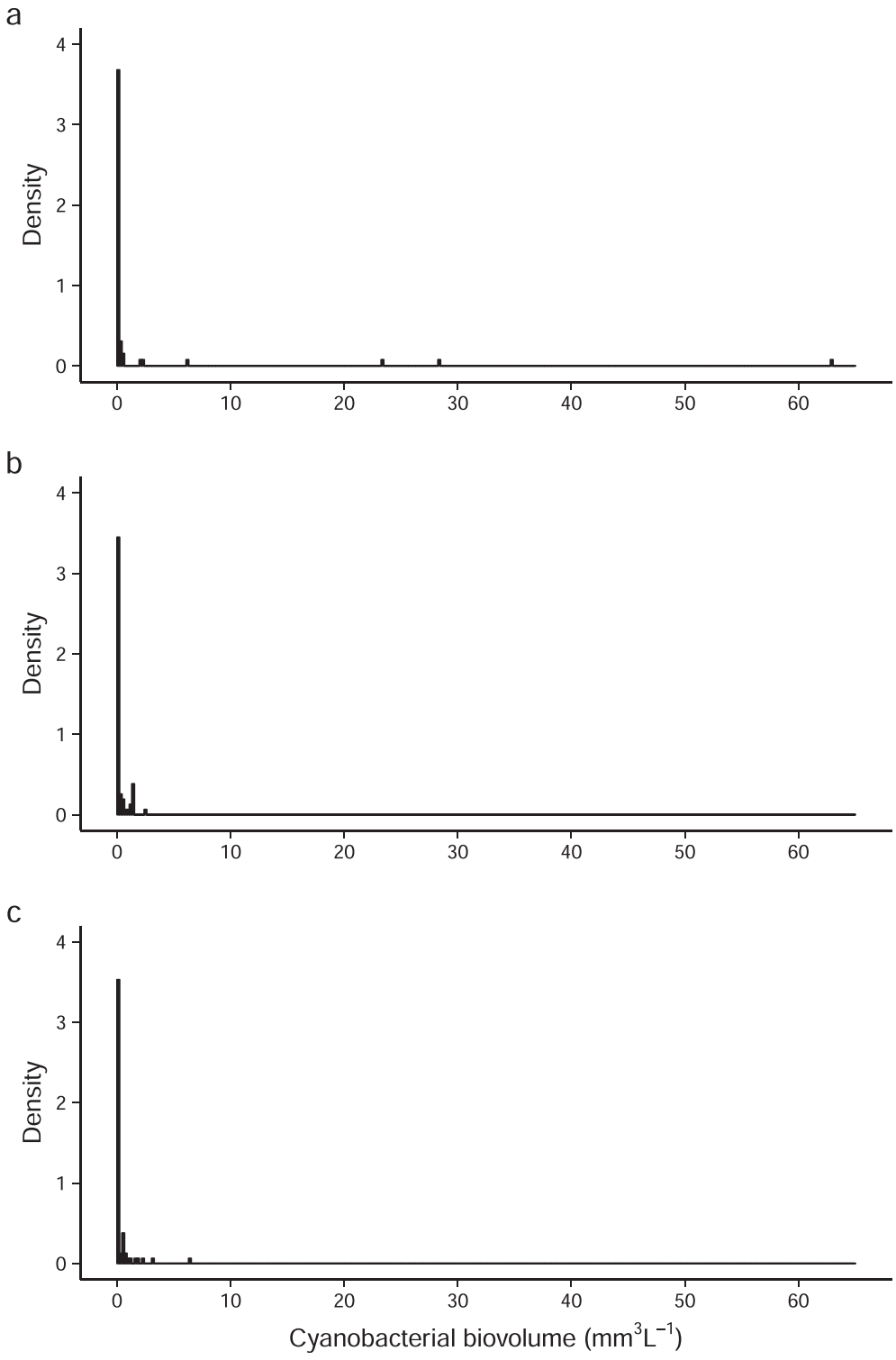

Fig. 3. Distribution of cyanobacterial biovolume at sampling sites 1 (a), 2 (b) and 3 (c) (site numbering as in Fig. 1).

at each site). For all sites combined (site means; n = 8 observations), it was $r^2 = 0.87$.

3.3. Predicting cyanobacterial exceedance probabilities and biovolume

The joint structuring effects of water temperature and conductivity on cyanobacteria were visualized by constructing contour plots of the probability of exceeding the three biovolume threshold levels (Fig. 7). As expected, the exceedance probability declined as threshold levels increased (Fig. 7). For the three tiers, the model predicted the highest probability at a temperature of 22.2 °C and at the lowest observed values for conductivity (~0.1 mS cm^{-1}) (Fig. 7). Overall, exceedance probabilities decreased

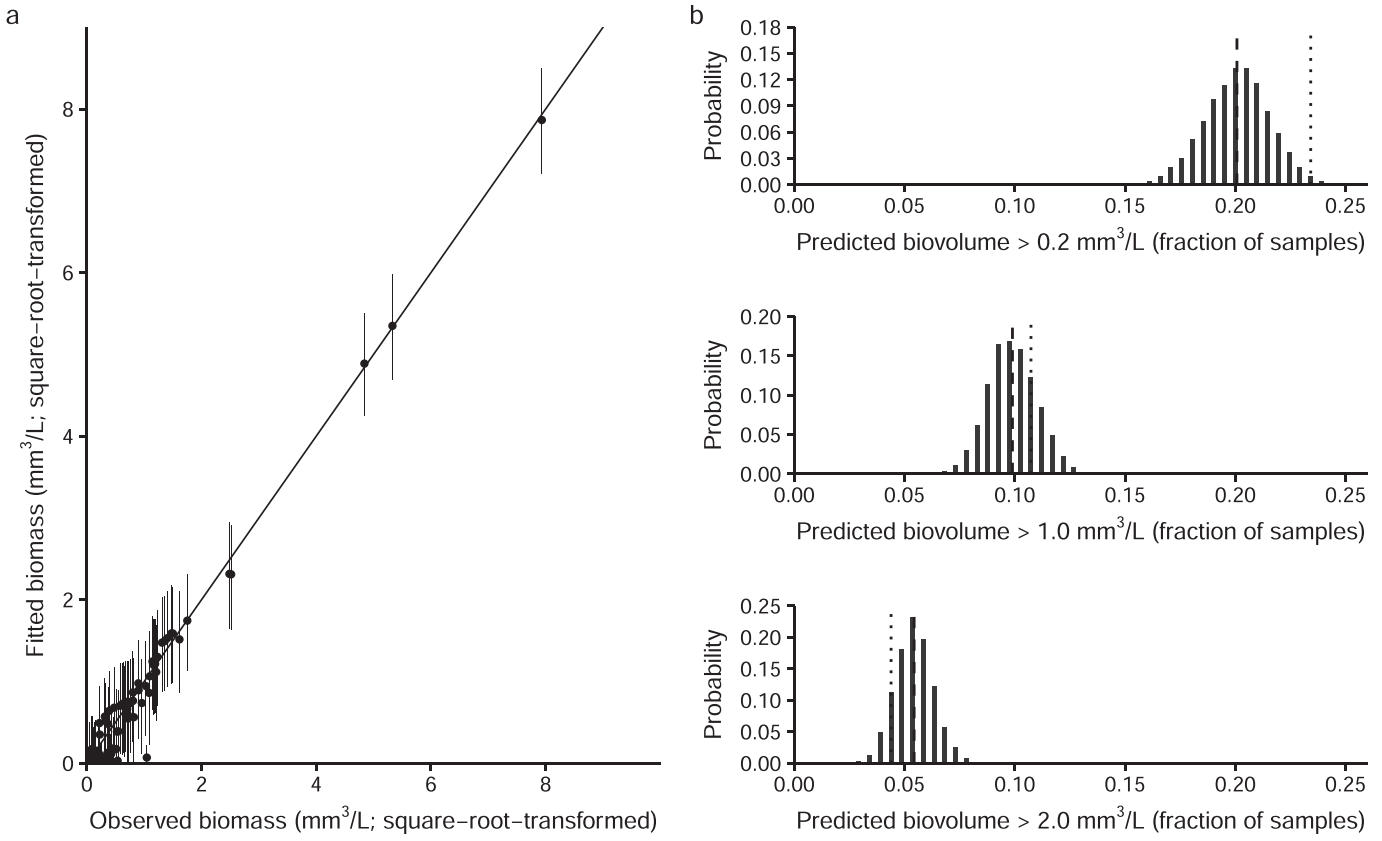

Fig. 4. (a) Observed versus fitted biomass (posterior distribution medians and 95% credible intervals), (b) Posterior predictive checks for the three risk thresholds (from top to bottom: > 0.2, >1.0 and >2.0 mm^3 L^{-1}), showing the distribution of the predicted proportion of samples exceeding each tier. Dashed vertical line: the mean of the predicted distribution, dotted vertical line: the observed proportion of samples exceeding the threshold.

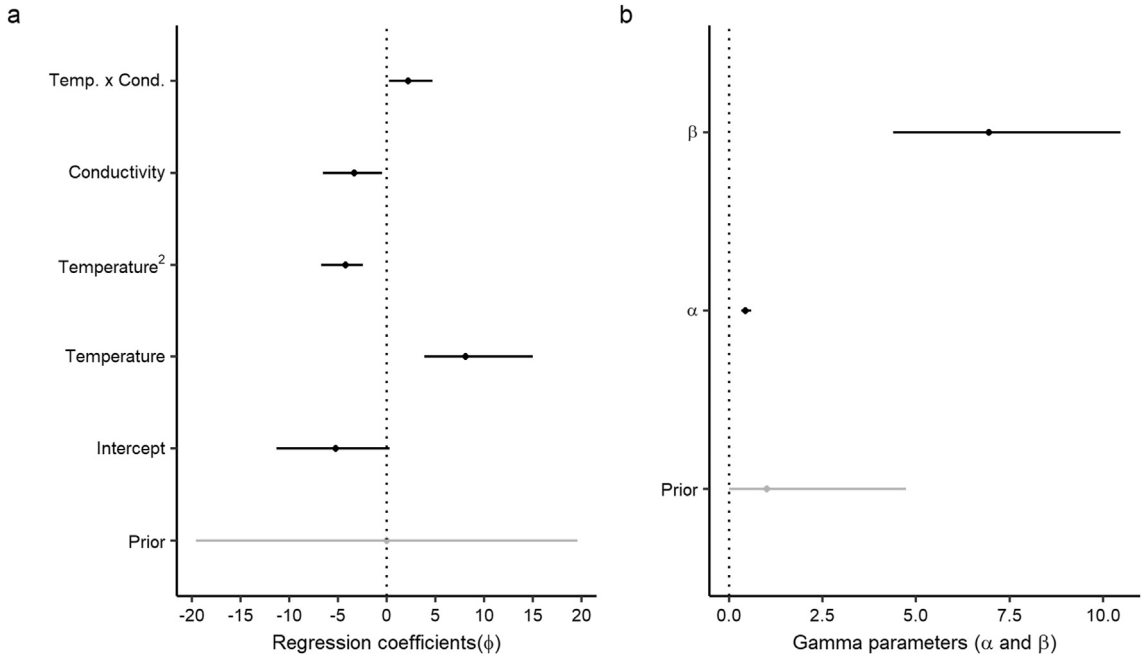

Fig. 5. CPG model parameters. Summary of the posterior distribution (medians and 95% credible intervals) of the regression coefficients (ϕ) and Gamma distribution shape and rate parameters (α and β).

with rising conductivity, and at the highest conductivity the temperature at which the maximum probability occurred was shifted upwards, to ~27 °C (Fig. 7). The nonlinearity in cyanobacterial responses to temperature and conductivity yielded a sharp increase

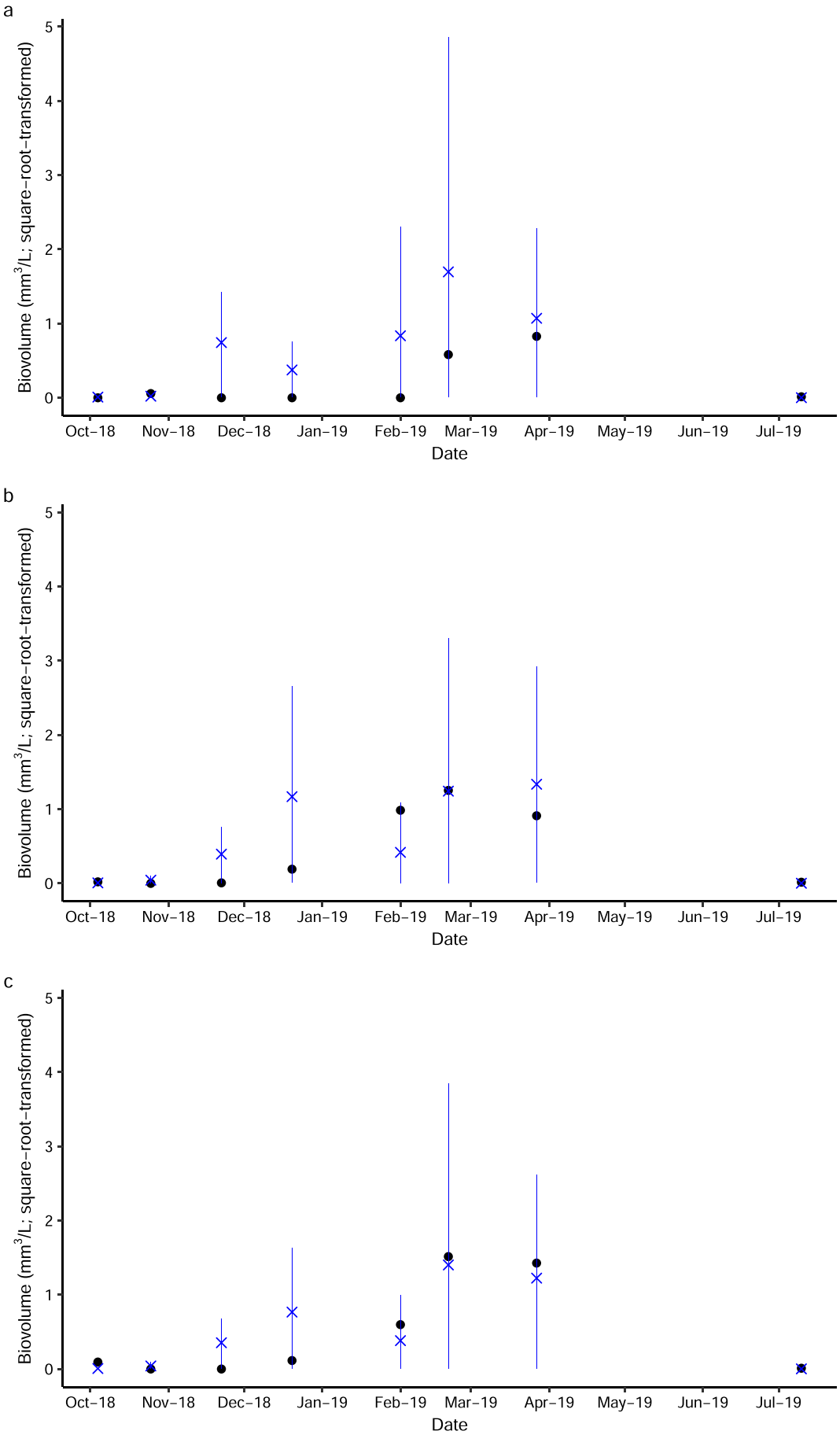

Fig. 6. Comparison of out-of-sample predicted values (mean and 95% credible interval; blue) and observed values (mean; black dots) for cyanobacterial biovolume at sampling sites 1 (a), 2 (b) and 3 (c) (site numbering as in Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

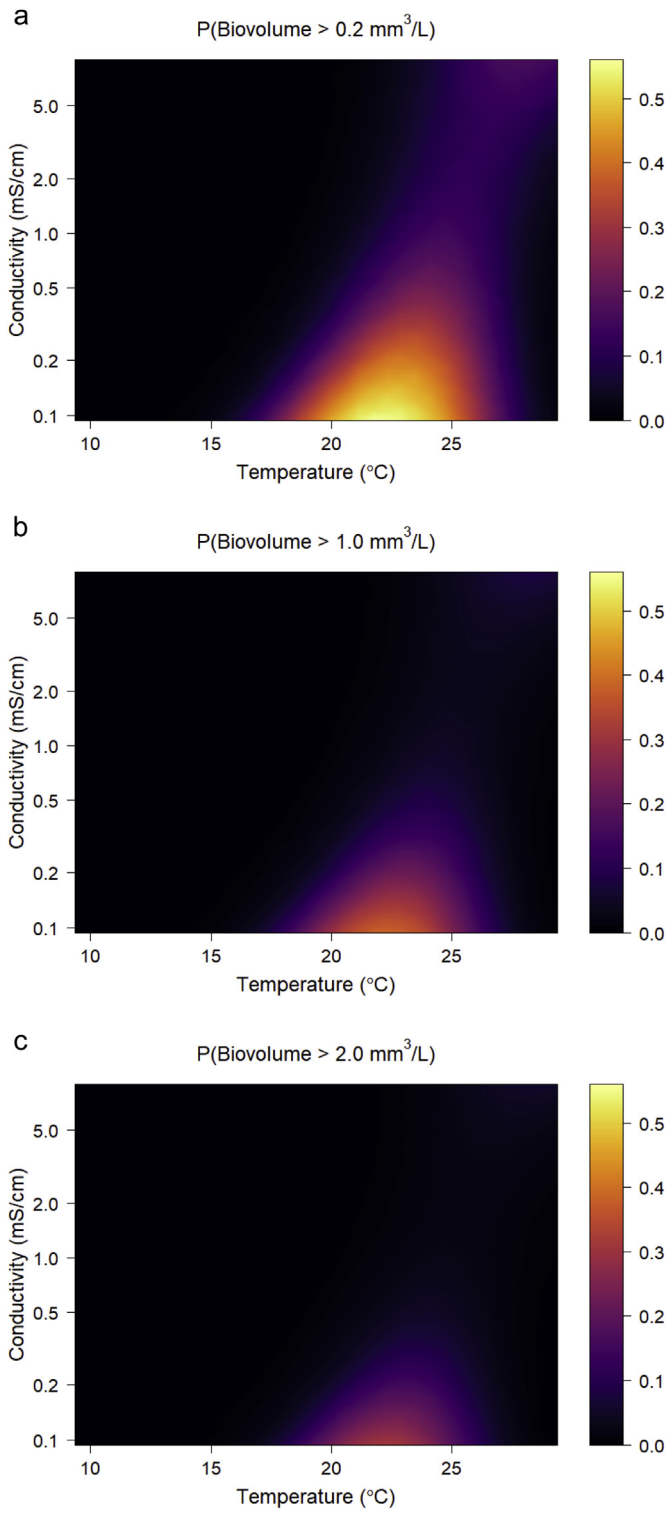

Fig. 7. Probability of exceeding three risk levels of exposure to cyanobacteria as a function of water temperature (°C) and conductivity (mS cm-1). (a) > 0.2 mm^3 L^{-1}; (b) > 1 mm^3 L^{-1}; (c) > 2 mm^3 L^{-1}.

in exceedance probabilities when conductivity falls below ~0.5 mS cm^{-1}. Temperature effects, such an increase from 20 to 25 °C, were most pronounced for the lowest threshold (Fig. 8). At 20 °C, cyanobacterial biovolume will most likely not reach any threshold when conductivity is greater than ~2 mS cm^{-1}, whereas at 25 °C even the highest tier has a substantial chance to be surpassed, even at high conductivity (Fig. 9).

4. Discussion

We presented a novel CPG approach to modeling cyanobacterial biovolume within a Bayesian probabilistic framework. This approach provides a full accounting of uncertainty in predictions, as it incorporates variation which originates from both sampling error and uncertainty in parameter values. Using two environmental variables, which are readily measured *in situ*, we were able to predict the probability of cyanobacterial biovolume exceeding three threshold values relevant to management. The risk of exposure expressed in probabilities can be a useful tool for early warning systems, particularly for coastal eutrophic environments that are susceptible to develop CyanoHABs.

CyanoHABs can occur within a very short period of time, shifting rapidly from undetected levels to peak biomass values (blooms). Our approach overcomes one of the problems inherent to statistical modeling of these organisms: the presence of a large proportion of zeros in data sets (Franks, 2018). For count data, such as cell abundance, hurdle models (e.g., Cha et al., 2014) assume that zeros occur from one process and the positive values from another, whereas zero-inflated models assume that "extra zeros" come from a second process (Cusack et al., 2015). However, for continuous data such as biovolume, this artificial categorization can impose an unwanted discontinuity in the data because in natural environments the same drivers often simultaneously influence absence (zeros) and presence (nonzero biomass) (Lecomte et al., 2013a). Contrary to most models for HABs, which use cell abundance as the response variable (e.g., Cha et al., 2014; Cusack et al., 2015), our approach is based on biovolume, which is the best indicator of phytoplankton biomass (Chorus and Cavalieri, 2000; Reynolds, 2006). Furthermore, the use of biovolume instead of other biological indicators can change model outputs and improve predictions of lake ecological status (Moe et al., 2016).

We found clear nonlinear, interactive effects of water temperature and conductivity on cyanobacterial biovolume. Growth of freshwater cyanobacteria is generally favored at higher temperatures, with well-defined thermal optima for growth at temperatures ranging between 20 and 30 °C (e.g., Fig. 5.3 in Reynolds, 2006), whereas high salinity tends to limit or inhibit cyanobacterial growth (Reynolds, 2006). The importance of these drivers for cyanobacterial blooms in estuaries has been well documented (temperature: Cha et al., 2014; Lehman et al., 2008; Robson and Hamilton, 2003; salinity: Hall et al., 2013; Robson and Hamilton, 2003; Taş et al., 2006). Models of cyanobacterial biomass have included interactions between environmental predictors (e.g., between temperature and nutrients in lentic water bodies; Taranu et al., 2012), but we found no previous examples of models including an interaction between temperature and conductivity. We suggest that inclusion of this interaction may lead to improvements in models for CyanoHABs in coastal environments.

Our model predicts that the probability of crossing the three threshold levels is greatest at 22.2 °C. Species from the genera *Microcystis* and *Dolichospermum* co-occur at our study site and frequently bloom together in the region (Aguilera et al., 2017; Haakonsson et al., 2017; O'Farrell et al., 2012) and elsewhere (Li et al., 2010; Te and Gin, 2011). In lentic ecosystems, mixed *Microcystis-Dolichospermum* blooms are commonly observed between 25 and 30 °C (Krienitz et al., 2002; O'Farrell et al., 2012; Te and Gin, 2011). However, in estuaries and coastal areas, such as the Baltic Sea and San Francisco Bay, cyanobacterial blooms are more frequent at temperatures nearer to 20 °C (Lehman et al., 2010; Pliński et al., 2007). A similar trend is found for coastal blooms in the Río de la Plata Estuary (5 out of 7 coastal blooms were

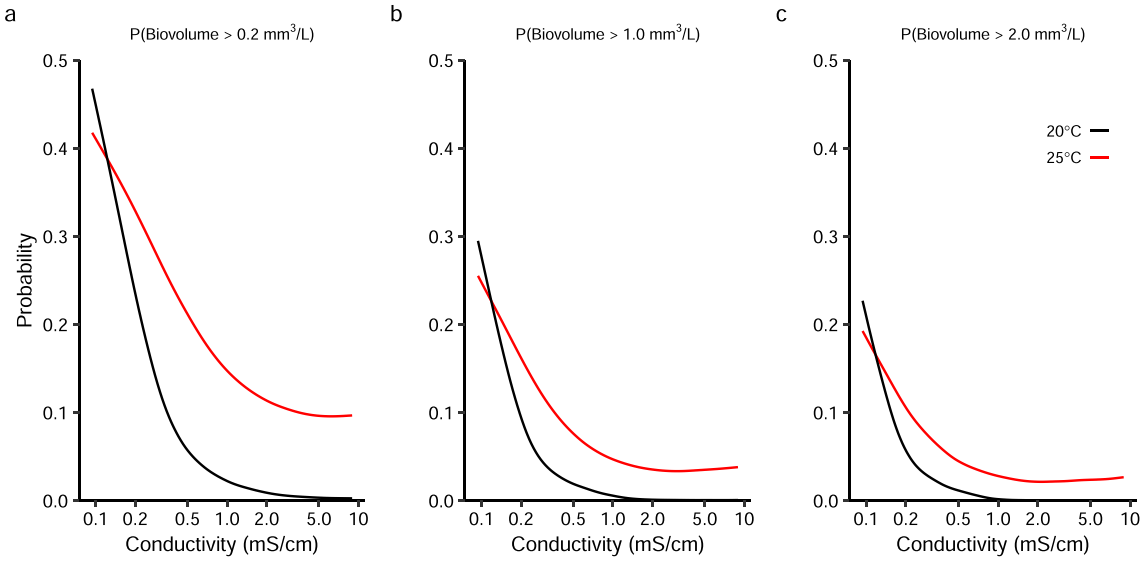

Fig. 8. Probability of exceeding three risk levels of exposure to cyanobacteria as a function of conductivity at two different temperatures: 20 °C (black curves) and 25 °C (red curves). (a) > 0.2 mm^3 L^{-1}; (b) > 1 mm^3 L^{-1}; (c) > 2 mm^3 L^{-1}. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

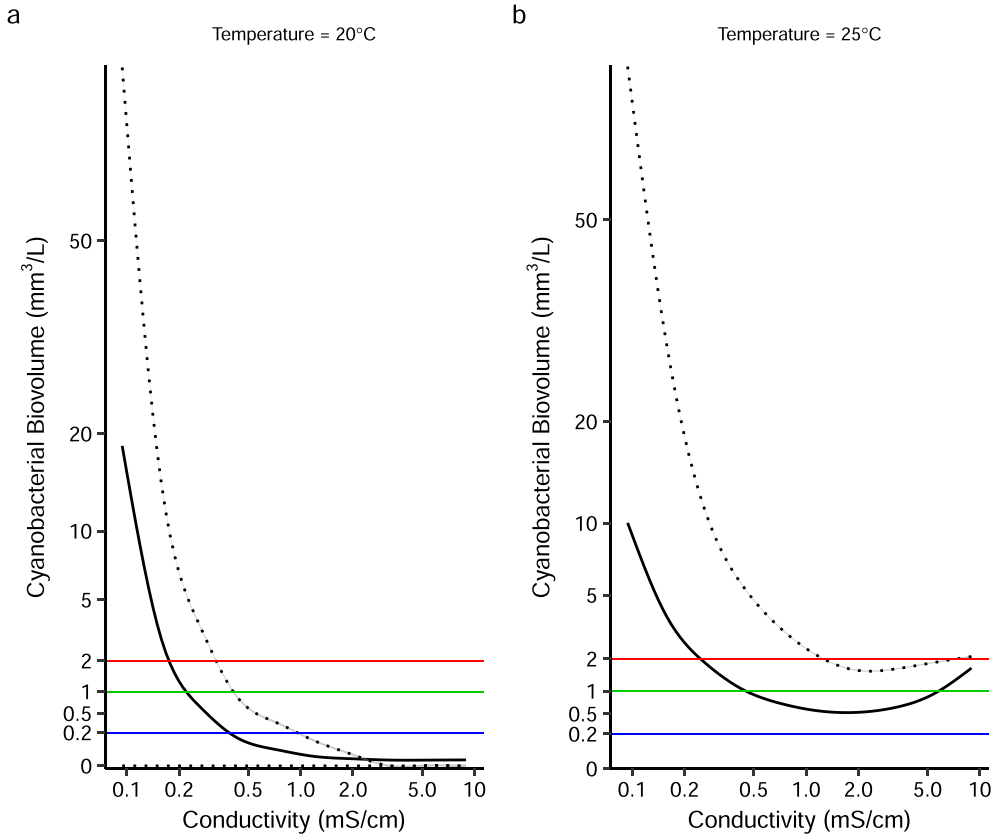

Fig. 9. Biovolume of cyanobacteria (mm^3 L^{-1}) as a function of conductivity at two water temperatures: (a) 20 °C; (b) 25 °C. Means (continuous black curves) and 25th and 75th percentiles (dotted curves) are shown. The horizontal lines represent the three risk levels of exposure to cyanobacteria (blue: 0.2 mm^3 L^{-1}; green: 1 mm^3 L^{-1}; red: 2 mm^3 L^{-1}). Note the square-root scale of the y axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

registered between 21 and 23 °C; Haakonsson et al., 2017). On the southern side of the estuary, cyanobacteria above WHO Guidance Level 1 (based on *Microcystis* cells) have been reported between 20 and 24 °C (Giannuzzi et al., 2012) (Supplementary Material Fig. S2). These findings, and our own results, allow us to highlight the potential of our new model to detect changes in cyanobacterial biovolume in temperature ranges commonly found for CyanoHABs in the studied region.

The two explanatory variables in our model, temperature and conductivity, measured *in situ*, are normally recorded in basic monitoring programs and can be obtained in real time, which potentially allows for instantaneous, on site determination of exceedance probabilities. This is a comparative advantage over models that require time-consuming laboratory measurements of predictors such as nutrients. Our model can therefore be a useful tool for early warning of CyanoHAB events in monitoring programs in eutrophic coastal waters. Furthermore, the conductivity values covered in the present study are in the salinity range of the oligohaline areas of estuaries (Elliott and McLusky, 2002), where blooms often occur (i.e.: Lehman et al., 2005; Taş et al., 2006). Finally, the increasing availability of freely available environmental data (e.g.: https://www.dinama.gub.uy/oan/datos-abiertos/https://www.dinama.gub.uy/oan/datos-abiertos/) should promote the use of this type of model.

Our results suggest that the probability of cyanobacterial bloom occurrence at the study site, and possibly in most coastal areas in the Río de la Plata Estuary, is strongly dependent on two physical variables directly related to climate. Annual precipitation (annual and seasonal means) is projected to increase in this region (IPCC, 2007), which should lower the conductivity and favor the proliferation of cyanobacteria (Reichwaldt and Ghadouani, 2012). When extreme rainfall occurs in the Río de la Plata basin during summer, blooms are likely to appear in the estuary (Kruk et al., 2019). Furthermore, freshwater discharge from two large rivers (Uruguay and Paraná) enters the estuary, resulting in reduced conductivity and increased temperature, and thus favoring cyanobacterial blooms (Kruk et al., 2019). Cyanobacterial biomass is also flushed from reservoirs upstream into the estuary where it continues to grow and proliferate (Kruk et al., 2019). Extreme rainfall events in the southern Río de la Plata basin are projected take place through the twenty-first century (Barros et al., 2013) and therefore cyanobacterial blooms are expected to increase in intensity and duration in the future (Haakonsson et al., 2017).

Despite the fact that nutrient concentrations can be major determinants of cyanobacterial growth, they may not be primary predictors of biomass in nutrient-rich estuaries (Robson and Hamilton, 2004; Taş et al., 2006). Although strong reductions in nutrient loads (phosphorus and nitrogen) may help control cyanobacterial growth (Boesch, 2002; Harke et al., 2016), implementing an on-site nutrient monitoring system with adequate instrumentation and spatio-temporal coverage remains a challenging task in a large ecosystem such as the Río de la Plata Estuary.

5. Conclusions

- The proposed CPG approach to modeling cyanobacterial biovolume successfully handled the high proportion of zeros and the extreme values typically found in CyanHAB samples and provided accurate predictions of cyanobacterial biomass and probabilities of exceedance for alert thresholds of interest to management.
- The key predictors in our model, temperature and conductivity, readily measured *in situ*, are normally recorded in basic monitoring programs and can be obtained in real time, which potentially allows for instantaneous prediction of cyanobacterial biomass.
- Our model has comparative advantages over models that require time-consuming laboratory measurements of predictors such as nutrients.
- In the context of climate change predictions, our model is a useful tool for early warning systems in coastal eutrophic environments that are susceptible to developing CyanoHABs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by ANII (FCE 6384 and POS_NAC_2016_1_130357), PEDECIBA (Uruguay), UTE (Uruguay) and ELAP (Canada). We thank Lorena Rodríguez-Gallego, Dermot Antoniades, and three anonymous reviewers for their helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watres.2020.115710.

References

Aguilera, A., Haakonsson, S., Martín, M.V., Echenique, R., Salerno, G., 2017. Bloom-forming cyanobacteria and cyanotoxins in Argentina: a growing health and environmental concern. Limnologica 69, 103–114. https://doi.org/10.1016/j.limno.2017.10.006.

Ahn, C.Y., Oh, H.M., Park, Y.S., 2011. Evaluation of environmental factors on cyanobacterial bloom in eutrophic reservoir using artificial neural networks. J. Phycol. 47, 495–504. https://doi.org/10.1111/j.1529-8817.2011.00990.x.

Barros, V.R., Garavaglia, C.R., Doyle, M.E., 2013. Twenty-first century projections of extreme precipitations in the Plata Basin. Int. J. River Basin Manag. 11, 373–387. https://doi.org/10.1080/15715124.2013.819358.

Beaulieu, M., Pick, F., Gregory-Eaves, I., 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes dataset. Limnol. Oceanogr. 58, 1736–1746. https://doi.org/10.4319/lo.2013.58.5.1736.

Boesch, D.F., 2002. Challenges and opportunities for science in reducing nutrient over-enrichment of coastal ecosystems. Estuaries 25, 886–900. https://doi.org/10.1007/BF02804914.

Bonilla, S., Haakonsson, S., Somma, A., Gravier, A., Britos, A., Vidal, L., De León, L., Brena, B., Pírez, M., Piccini, C., Martínez de la Escalera, G., Chalar, G., González-Piana, M., Martigani, F., Aubriot, L., 2015. Cianobacterias y cianotoxinas en ecosistemas límnicos de Uruguay [Cyanobacteria and cyanotoxins in freshwaters of Uruguay]. INNOTEC 10, 9–22.

Bonilla, S., Pick, F.R., 2017. Freshwater bloom-forming cyanobacteria and anthropogenic change. Limnol. Oceanogr. E-Lecture 7, 1–62. https://doi.org/10.1002/loe2.10006.

Brena, B.M., Diaz, L., Sienra, D., Ferrari, G., Ferraz, N., Hellman, U., Gonzalez-Sapienza, G., Last, J.A., 2006. ITREOH building of regional capacity to monitor recreational water: development of a non-comercial microcystin ELISA and its impact on public health policy. UC Davis Previously Publ. Work. UC Davis 12, 377–385. https://doi.org/10.1179/oeh.2006.12.4.377.

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 Resour. Res. 50, 2518–2532. https://doi.org/10.1002/2013WR014372.

Carey, C.C., Ibelings, B.W., Hoffmann, E.P., 2012. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. Water Res. 46 (5), 1394–1407. https://doi.org/10.1016/j.watres.2011.12.016.

Cha, Y.K., Cho, K.H., Lee, H., Kang, T., Kim, J.H., 2017. The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. Water Res. 124, 11–19. https://doi.org/10.1016/j.watres.2017.07.040.

Chorus, I., Bartram, J., 1999. Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management. E & FN Spon, London.

Chorus, I., Cavalieri, M., 2000. Chapter 10. Cyanobacteria and algae. In: Monitoring Bathing Waters - A Practical Guide to the Design and Implementation of Assessments and Monitoring Programmes, pp. 205–258.

Cusack, C., Mouriño, H., Moita, M.T., Silke, J., 2015. Modelling Pseudo-nitzschia events off southwest Ireland. J. Sea Res. 105, 30–41. https://doi.org/10.1016/j.seares.2015.06.012.

Davis, J.R., Koop, K., 2006. Eutrophication in Australian rivers, reservoirs and estuaries - a southern hemisphere perspective on the science and its implications. Hydrobiologia 559, 23–76. https://doi.org/10.1007/s10750-005-4429-2.

De León, L., Yunes, J., 2001. First report of a Microcystis aeruginosa toxic bloom in La Plata River. Environ. Toxicol. Water Qual. 16, 110–112.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., Pitts, K.L., 2009. Eutrophication of U.S. Freshwaters: analysis of potential economic damages. Environ. Sci. Technol. 43, 12–19. https://doi.org/10.1021/es801217q.

Dolman, A.M., Rücker, J., Pick, F.R., Fastner, J., Rohrlack, T., Mischke, U., Wiedner, C., 2012. Cyanobacteria and cyanotoxins: the influence of nitrogen versus phosphorus. PloS One 7, e38757. https://doi.org/10.1371/journal.pone.0038757.

Downing, J.A., Watson, S.B., McCauley, E., 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. 58, 1905–1908. https://doi.org/10.1139/f01-143.
Elliott, M., McLusky, D.S., 2002. The need for definitions in understanding estuaries. Estuar. Coast Shelf Sci. 55, 815–827. https://doi.org/10.1006/ecss.2002.1031.
Elliott, J.A., 2010. The seasonal sensitivity of Cyanobacteria and other phytoplankton to changes in flushing rate and water temperature. Global Change Biol. 16, 864–876. https://doi.org/10.1111/j.1365-2486.2009.01998.x.
Engström-Öst, J., Repka, S., Mikkonen, M., 2011. Interactions between plankton and cyanobacterium *Anabaena* with focus on salinity, growth and toxin production. Harmful Algae 10, 530–535. https://doi.org/10.1016/j.hal.2011.04.002.
Fletcher, D., MacKenzie, D., Villouta, E., 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environ. Ecol. Stat. V12, 45–54.
Foster, S.D., Bravington, M.V., 2013. A Poisson–Gamma model for analysis of ecological non-negative continuous data. Environ. Ecol. Stat. 20, 533–552. https://doi.org/10.1007/s10651-012-0233-0.
Franks, P.J.S., 2018. Recent advances in modelling of harmful algal blooms. In: Glibert, P.M., Berdalet, E., Burford, M., Pitcher, G., Zhou, M. (Eds.), Global Ecology and Oceanography of Harmful Algal Blooms. Ecological Studies (Analysis and Synthesis), 232. Springer, Cham, pp. 359–377. https://doi.org/10.1007/978-3-319-70069-4_19.
Gelman, A., Meng, X.-L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sin. 6, 733–807.
Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., 2014. Bayesian Data Analysis, third ed. CRC Press.
Ghaffar, S., Stevenson, R.J., 2016. Cyanobacteria dominance in lakes and evaluation of its predictors: a study of Southern Appalachians ecoregion, USA. MATEC Web Conf 60, 1–5. https://doi.org/10.1051/matecconf/20166002001.
Giannuzzi, L., Carvajal, G., Corradini, M.G., Araujo Andrade, C., Echenique, R., Andrinolo, D., 2012. Occurrence of toxic cyanobacterial blooms in Río de la Plata Estuary, Argentina: field study and data analysis. J. Toxicol. 2012, 1–15. https://doi.org/10.1155/2012/373618.
Giannuzzi, L., Colombi, A., Pruyas, T., Aún, A., Rujana, M., Falcione, M., Zubieta, J., 2009. Cianobacterias y Cianotoxinas: Identificación, Toxicología Monitoreo y Evaluación de Riesgo. Moglia Impresiones, Corrientes.
Glibert, P.M., Allen, J.I., Bouwman, A.F., Brown, C.W., Flynn, K.J., Lewitus, A.J., Madden, C.J., 2010. Modeling of HABs and eutrophication: status, advances, challenges. J. Mar. Syst. 83, 262–275. https://doi.org/10.1016/j.jmarsys.2010.05.004.
Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Stow, C.A., Van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae 54, 87–97. https://doi.org/10.1016/j.hal.2016.01.010.
Gómez, N., 2014. Phytoplankton of the Río de la Plata estuary. Adv. Limnol. 65, 167–182. https://doi.org/10.1127/1612-166X/2014/0065-040.
Haakonsson, S., Rodríguez-Gallego, L., Somma, A., Bonilla, S., 2017. Temperature and precipitation shape the distribution of harmful cyanobacteria in subtropical lotic and lentic ecosystems. Sci. Total Environ. 609, 1132–1139. https://doi.org/10.1016/j.scitotenv.2017.07.067.
Hall, N.S., Paerl, H.W., Peierls, B.L., Whipple, A.C., Rossignol, K.L., 2013. Effects of climatic variability on phytoplankton community structure and bloom development in the eutrophic, microtidal, New River Estuary, North Carolina, USA. Estuar. Coast Shelf Sci. 117, 70–82. https://doi.org/10.1016/j.ecss.2012.10.004.
Hamilton, G., McVinish, R., Mengersen, K., 2009. Bayesian model averaging for harmful algal bloom prediction. Ecol. Appl. 19, 1805–1814. https://doi.org/10.1890/08-1843.1.
Harke, M.J., Steffen, M.M., Gobler, C.J., Otten, T.G., Wilhelm, S.W., Wood, S.A., Paerl, H.W., 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium. *Microcystis* spp. Harmful Algae 54, 4–20. https://doi.org/10.1016/j.hal.2015.12.007.
Hillebrand, H., Durselen, C., Kirschtel, D., Pollingher, U., Zohary, T., 1999. Biovolume calculation for pelagic and benthic microalgae. J. Phycol. 35, 403–424.
Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B., Verspagen, J., Visser, P., 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. 16, 471–483. https://doi.org/10.1038/s41579-018-0040-1.
Ibelings, B.W., Backer, L.C., Kardinaal, W.E.A., Chorus, I., 2015. Current approaches to cyanotoxin risk assessment and risk management around the globe. Harmful Algae 49, 63–74. https://doi.org/10.1016/j.hal.2014.10.002.
IPCC, 2007. Climate Change 2007: the Physical Science Basis. A Report of Working Group 1 of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p. 996.
Johnson, S., Fielding, F., Hamilton, G., Mengersen, K., 2010. An Integrated Bayesian Network approach to Lyngbya majuscula bloom initiation. Mar. Environ. Res. 69, 27–37. https://doi.org/10.1016/j.marenvres.2009.07.004.
Kosten, S., Huszar, V.L.M., Bécares, E., Costa, L.S., Donk, E., Hansson, L.-A., Jeppesen, E., Kruk, C., Lacerot, G., Mazzeo, N., Meester, L., Moss, B., Lürling, M., Nõges, T., Romo, S., Scheffer, M., 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Global Change Biol. 18, 118–126. https://doi.org/10.1111/j.1365-2486.2011.02488.x.
Krienitz, L., Ballot, A., Claudia, W., Kotut, K., Codd, G.A., Pflugmacher, S., 2002. Cyanotoxin producing bloom of *Anabaena flos-aquae, Anabaena discoidea* and *microcystis aeruginosa* (Cyanobacteria) in Nyanza Gulf of Lake Victoria, Kenya. J. Appl. Bot. - Angew. Bot. 76, 179–183.
Kruk, C., Martínez, A., Martínez de la Escalera, G., Trinchin, R., Manta, G., Segura, A.M., Piccini, C., Brena, B.M., Fabiano, G., Pirez, M., Gabito, L., Alcantara, I., Yannicelli, B., 2019. Floración excepcional de cianobacterias tóxicas en la costa de Uruguay, verano 2019 Exceptional [Exceptional bloom of toxic cyanobacteria on the Uruguayan coast, summer 2019]. INNOTEC 18, 36–68. https://doi.org/10.26461/18.06.
Kruk, C., Segura, A., Nogueira, L., Carballo, C., Martínez de la Escalera, G., Calliari, D., Ferrari, G., Simoens, M., Cea, J., Alcantara, I., Vico, P., Miguez, D., Piccini, C., 2015. Herramientas para el monitoreo y sistema de alerta de floraciones de cianobacterias nocivas: Río Uruguay y Río de la Plata [Monitoring tools and early warning system for harmful cyanobacterial blooms: Río Uruguay and Río de la Plata]. INNOTEC 10, 23–39.
Lancelot, C., Muylaert, K., 2011. Trends in Estuarine Phytoplankton Ecology, Treatise on Estuarine and Coastal Science: Functioning Ecosystems at the Land-Ocean Interface. Elsevier Inc., Amsterdam https://doi.org/10.1016/B978-0-12-374711-2.00703-8.
Lecomte, J.B., Benoît, H.P., Ancelet, S., Etienne, M.P., Bel, L., Parent, E., 2013a. Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. Methods Ecol. Evol 4, 1159–1166. https://doi.org/10.1111/2041-210X.12122.
Lecomte, J.B., Benoît, H.P., Etienne, M.P., Bel, L., Parent, E., 2013b. Modeling the habitat associations and spatial distribution of benthic macroinvertebrates: a hierarchical Bayesian model for zero-inflated biomass data. Ecol. Model. 265, 74–84. https://doi.org/10.1016/J.ECOLMODEL.2013.06.017.
Lehman, P.W., Boyer, G., Hall, C., Waller, S., Gehrts, K., 2005. Distribution and toxicity of a new colonial *Microcystis aeruginosa* bloom in the San Francisco Bay Estuary, California. Hydrobiologia 541, 87–99. https://doi.org/10.1007/s10750-004-4670-0.
Lehman, P.W., Boyer, G., Satchwell, M., Waller, S., 2008. The influence of environmental conditions on the seasonal variation of *Microcystis* cell density and microcystins concentration in San Francisco Estuary. Hydrobiologia 600, 187–204. https://doi.org/10.1007/s10750-007-9231-x.
Lehman, P.W., Teh, S.J., Boyer, G.L., Nobriga, M.L., Bass, E., Hogle, C., 2010. Initial impacts of *Microcystis aeruginosa* blooms on the aquatic food web in the San Francisco Estuary. Hydrobiologia 637, 229–248. https://doi.org/10.1007/s10750-009-9999-y.
Li, Z., Yu, J., Yang, M., Zhang, J., Burch, M.D., Han, W., 2010. Cyanobacterial population and harmful metabolites dynamics during a bloom in Yanghe Reservoir, North China. Harmful Algae 9, 481–488. https://doi.org/10.1016/j.hal.2010.03.003.
Martínez de la Escalera, G., Kruk, C., Segura, A.M., Nogueira, L., Alcántara, I., Piccini, C., 2017. Dynamics of toxic genotypes of *Microcystis aeruginosa* complex (MAC) through a wide freshwater to marine environmental gradient. Harmful Algae 62, 73–83. https://doi.org/10.1016/j.hal.2016.11.012.
Moe, S.J., Haande, S., Couture, R.M., 2016. Climate change, cyanobacteria blooms and ecological status of lakes: a Bayesian network approach. Ecol. Model. 337, 330–347. https://doi.org/10.1016/j.ecolmodel.2016.07.004.
Nagy, G.J., Gómez-Erache, M., López, C.H., Perdomo, A.C., 2002. Distribution patterns of nutrients and symptoms of eutrophication in the Río de la Plata River Estuary System. Hydrobiologia 475–476, 125–139. https://doi.org/10.1023/A:1020300906000.
O'Farrell, I., Bordet, F., Chaparro, G., 2012. Bloom forming cyanobacterial complexes co-occurring in a subtropical large reservoir: validation of dominant eco-strategies. Hydrobiologia 698, 175–190. https://doi.org/10.1007/s10750-012-1102-4.
O'Neil, J.M., Davis, T.W., Burford, M.A., Gobler, C.J., 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae 14, 313–334. https://doi.org/10.1016/j.hal.2011.10.027.
Obenour, D., Gronewold, A., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resour. Res. 50, 7847–7860. https://doi.org/10.1002/2014WR015616.
Oliver, R.L., Hamilton, D.P., Brookes, J.D., Ganf, G.G., 2012. Physiology, blooms and prediction of planktonic cyanobacteria. In: Whitton, B.A. (Ed.), Ecology of Cyanobacteria II. Springer, Dordrecht, pp. 155–194. https://doi.org/10.1007/978-94-007-3855-3.
Olli, K., Paerl, H.W., Klais, R., 2015. Diversity of coastal phytoplankton assemblages - cross ecosystem comparison. Estuar. Coast Shelf Sci. 162, 1–9. https://doi.org/10.1016/j.ecss.2015.03.015.
Paerl, H.W., Huisman, J., 2008. Blooms like it hot. Sci. Magna 320, 57–58.
Pennington, M., 1996. Estimating the mean and variance from highly skewed marine data. Fish. Bull. 94, 498–505.
Persaud, A.D., Paterson, A.M., Dillon, P.J., Winter, J.G., Palmer, M., Somers, K.M., 2015. Forecasting cyanobacteria dominance in Canadian temperate lakes. J. Environ. Manag. 151, 343–352. https://doi.org/10.1016/j.jenvman.2015.01.009.
Pliński, M., Mazur-Marzec, H., Jóźwiak, T., Kobos, J., 2007. The potential causes of cyanobacterial blooms in Baltic Sea estuaries. Oceanol. Hydrobiol. Stud. 36, 125–137. https://doi.org/10.2478/v10009-007-0001-x.
Plummer, M., 2014. Rjags: Bayesian Graphical Models Using MCMC. R Packages Version 4–8.
Plummer, M., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.
Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R. News 6, 7–11.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. http://www.r-project.org/.

Reichwaldt, E.S., Ghadouani, A., 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. Water Res. 46, 1372–1393. https://doi.org/10.1016/j.watres.2011.11.052.

Reynolds, C.S., 2006. Ecology of Phytoplankton. Cambridge University Press, Cambridge.

Rigosi, A., Hanson, P.C., Hamilton, D.P., Hipsey, M.R., Rusak, J.A., Bois, J., Sparber, K., Chorus, I., Watkinson, A.J., Qin, B., Kim, B., Brookes, J.D., 2015. Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. Ecol. Appl. 25, 186–199.

Robson, B.J., Hamilton, D.P., 2004. Three-dimensional modelling of a *Microcystis* bloom event in the swan river estuary, western Australia. Ecol. Model. 174, 203–222. https://doi.org/10.1016/j.ecolmodel.2004.01.006.

Robson, B.J., Hamilton, D.P., 2003. Summer flow event induces a cyanobacterial bloom in a seasonal Western Australian estuary. Mar. Freshw. Res. 54, 139. https://doi.org/10.1071/mf02090.

Sathicq, M.B., Gómez, N., Andrinolo, D., Sedan, D., Donadelli, J.L., 2014. Temporal distribution of cyanobacteria in the coast of a shallow temperate estuary (Río de la Plata): some implications for its monitoring. Environ. Monit. Assess. 186, 7115–7125. https://doi.org/10.1007/s10661-014-3914-3.

Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310. https://doi.org/10.2139/ssrn.1351252.

Skrondal, A., 2009. Prediction in multilevel generalized linear models. J. Roy. Stat. Soc. A 172, 659–687.

Smith, V.H., 1986. Light and nutrient effects on the relative biomass of blue-green algae in lake phytoplankton. Can. J. Fish. Aquat. Sci. 43, 148–153. https://doi.org/10.1139/f86-016.

Smith, V.H., Schindler, D.W., 2009. Eutrophication science: where do we go from here? Trends Ecol. Evol. 24, 201–207. https://doi.org/10.1016/j.tree.2008.11.009.

Smyth, G.K., 1996. Regression analysis of quantity data with exact zeroes. In: Proceedings of the Second Australia-Japan Workshop on Stochastic Models in Engineering, Technology and Management. Technology Management Centre, University of Queensland, pp. 572–580.

Sournia, A., 1978. Phytoplankton manual. Monogr. Oceanogr. Methodol. 6, 337.

Steffensen, D.A., 2008. Economic cost of cyanobacterial blooms. In: Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. Springer New York, New York, NY, pp. 855–865. https://doi.org/10.1007/978-0-387-75865-7_37.

Taranu, Z.E., Gregory-Eaves, I., Steele, R.J., Beaulieu, M., Legendre, P., 2017. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: a new framework for modelling an important health risk factor. Global Ecol. Biogeogr. 26, 625–637. https://doi.org/10.1111/geb.12569.

Taranu, Z.E., Zurawell, R.W., Pick, F., Gregory-Eaves, I., 2012. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. Global Change Biol. 18, 3477–3490. https://doi.org/10.1111/gcb.12015.

Taş, S., Okuş, E., Aslan-Yilmaz, A., 2006. The blooms of a cyanobacterium, Microcystis cf. aeruginosa in a severely polluted estuary, the Golden Horn, Turkey. Estuar. Coast Shelf Sci. 68, 593–599. https://doi.org/10.1016/j.ecss.2006.02.025.

Te, S.H., Gin, K.Y.H., 2011. The dynamics of cyanobacteria and microcystin production in a tropical reservoir of Singapore. Harmful Algae 10, 319–329. https://doi.org/10.1016/j.hal.2010.11.006.

Utermöhl, H., 1958. Zur vervollkommnung der quantitativen phytoplankton-methodik. Mitt. Int. Ver. Theor. Angew. Limnol. 9, 1–38.

Wagner, C., Adrian, R., 2009. Cyanobacteria dominance: quantifying the effects of climate change. Limnol. Oceanogr. 54, 2460–2468. https://doi.org/10.4319/lo.2009.54.6_part_2.2460.

Withers, C., Nadarajah, S., 2011. On the compound Poisson-gamma distribution. Kybernetika 47, 15–37.

Yang, J., Lv, H., Yang, J., Liu, L., Yu, X., Chen, H., 2016. Decline in water level boosts cyanobacteria dominance in subtropical reservoirs. Sci. Total Environ. 557 (558), 445–452. https://doi.org/10.1016/j.scitotenv.2016.03.094.