Water Research 186 (2020) 116356

Contents lists available at ScienceDirect

Water Research

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

Using Bayesian hierarchical modelling to capture cyanobacteria dynamics in Northern European lakes

Nikolaos K. Mellios[a], S. Jannicke Moe[b], Chrysi Laspidou[a,*]

[a] Department of Civil Engineering, University of Thessaly, 38334 Volos, Greece
[b] Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway

ARTICLE INFO

Article history:
Received 18 May 2020
Revised 17 August 2020
Accepted 28 August 2020
Available online 28 August 2020

Keywords:
Cyanobacteria
Nutrients
Water framework directive
Bayesian Hierarchical modelling
Eutrophication lake management
WHO risk levels

ABSTRACT

Cyanobacteria blooms in lakes and reservoirs currently threaten water security and affect the ecosystem services provided by these freshwater ecosystems, such as drinking water and recreational use. Climate change is expected to further exacerbate the situation in the future because of higher temperatures, extended droughts and nutrient enrichment, due to urbanisation and intensified agriculture. Nutrients are considered critical for the deterioration of water quality in lakes and reservoirs and responsible for the widespread increase in cyanobacterial blooms. We model the response of cyanobacteria abundance to variations in lake Total Phosphorus (TP) and Total Nitrogen (TN) concentrations, using a data set from 822 Northern European lakes. We divide lakes in ten groups based on their physico-chemical characteristics, following a modified lake typology defined for the Water Framework Directive (WFD). This classification is used in a Bayesian hierarchical linear model which employs a probabilistic approach, transforming uncertainty into probability thresholds. The hierarchical model is used to calculate probabilities of cyanobacterial concentrations exceeding risk levels for human health associated with the use of lakes for recreational activities, as defined by the World Health Organization (WHO). Different TN and TP concentration combinations result in variable probabilities to exceed pre-set thresholds. Our objective is to support lake managers in estimating acceptable nutrient concentrations and allow them to identify actions that would achieve compliance of cyanobacterial abundance risk levels with a given confidence level.

1. Introduction

Freshwater is inextricably linked to human well-being and socio-economic development, while this dependence is a key condition for the sustainable management of freshwater resources. As the planet's population increases and, as a consequence, urbanisation and agriculture intensify, freshwater security is threatened by the growing demand for food production, electrical power generation, industrial processes and human consumption. On top of that, water quality generally suffers from continuous degradation in many regions, and as a result, freshwater ecosystems often become inhospitable habitats for living organisms (UNEP, 2016). This trend is expected to worsen in the near future and next generations are likely to face significant adverse impacts on water quantity and quality, especially under the threat of climate change. In recent years, global concern has evolved into specific action plans for water management; the United Nations released an agenda defining sustainable development goals (SDGs), where water management holds a prominent position calling for actions by all countries to increase the access to clean drinking water and sanitation (SDG 6) and to conserve and use oceans seas and marine resources sustainably (SDG 14) (UN 2019). Furthermore, since 2000, the European Water Framework Directive (WFD) has transformed water management in Europe, by bringing aquatic ecology to the forefront of decisions (Hering et al., 2010). Traditionally, the only common biological indicator of lake quality assessment and management was Chlorophyl-a (Chl-a), but following the implementation of the WFD, cyanobacteria abundance has become an additional indicator required for assessment of ecological status for European lakes (Birk et al., 2012).

Harmful cyanobacterial blooms pose a serious risk to freshwater quality, affecting human and animal health. Due to the toxins released by many bloom-forming species, water becomes inappropriate to serve human needs such as drinking water, fisheries

* Corresponding author.
E-mail addresses: nmellios@uth.gr (N.K. Mellios), jmo@niva.no (S.J. Moe), laspidou@uth.gr (C. Laspidou).

https://doi.org/10.1016/j.watres.2020.116356

and recreation (Charmichael et al., 2016; Lévesque et al., 2014; Ibelings et al., 2016). Scientific research has paid considerable attention to predicting the frequency and extent of cyanobacterial bloom events, suggesting possible interventions to mitigate these phenomena (Jewett et al., 2008; Tromas et al., 2017). However, predicting cyanobacterial abundance remains a challenge: even though there is good understanding of the key factors that drive and influence cyanobacterial dynamics, there is still high variability, making it difficult to accurately predict the abundance. In addition, availability of data exhibits great variation among freshwater ecosystems, making it difficult to come up with robust methodologies that would be applicable to a wide range of lake types (Richardson et al., 2018).

A common practice for bridging the gap of insufficient data in lake ecosystems is to "borrow" data from lakes with similar characteristics and in this way expand the sample size towards strengthening statistical analyses. However, when predictive models are applied to lakes categorized to groups following the assumption of homogeneity, the results usually fail to prove realistic, since homogeneity within a lake group is a weak assumption (Malve and Qian, 2006). Beaulieu et al. (2013) used a 1000 lake dataset containing data from lakes across the United States and implemented multiple linear regression analyses to predict cyanobacterial biomass on the whole dataset and on subsets of lake type according to depth and to whether the ecosystem is natural or a reservoir. The findings of this analysis indicated that predictions improved when lakes were categorized to groups; however, the overall low predictive strength advocates that the grouping assumption alone lacks satisfactory results. In another study conducted by Richardson et al. (2018), the response of cyanobacteria to multiple stressors by using linear regression mixed effect models varied greatly with lake type, resulting in the conclusion that a "one-size fits-all" approach is inappropriate towards understanding and managing the risks of harmful algal blooms.

Carvalho et al. (2013), used quantile regression modelling to quantify the relationship between TP concentrations and cyanobacteria, using a data set from 800 European lakes. The analysis showed that TP cannot be singled out as the dominant factor regarding cyanobacteria concentrations in lakes; rather, TP quantile modelling can be used to define the lake maximum cyanobacteria abundance, but only in relation to TP. Even though it is widely recognized that total nitrogen (TN) also plays a key role in cyanobacteria, previous modelling efforts of cyanobacteria in large datasets focus only on TP in their models (Richardson et al., 2018; Carvalho et al., 2013; Obenour et al., 2014). In their work based on mesocosm experiments, Richardson et al. (2019) include both TN and TP but only in combination, not separately, so the interaction with cyanobacteria cannot be analysed. Our work addresses this gap, as it uses both TN and TP separately as predictors for cyanobacteria.

Bayesian hierarchical models can combine prior and data-driven knowledge both from multiple groups of lakes and from lakes of the same group in order to make predictions for a single lake belonging to a specific group. In other words, the hierarchical approach moves one step further from the classical "grouping" approach by considering the effects of the ensemble of lakes on predictions. The Bayesian modelling framework, which is based on probability distributions is very suitable for the analysis of cyanobacteria blooms, as they are rare events with high uncertainty. The method has been used extensively in the past with convincing results (e.g., Malve and Qian, 2006; Shimoda and Archonditsis, 2015; Shimoda et al., 2016; Cheng et al., 2009; Obenour et al., 2014; Stow et al., 2014); however, the method has not been used for prediction of cyanobacteria abundance. Specifically, Malve and Qian (2006) have developed a similar modelling framework, using TN and TP as predictors, but they only predicted Chl-a. This research is to our knowledge novel because it models cyanobacteria using a Bayesian hierarchical model with both TN and TP as predictors; herein, we build upon the work of Malve and colleagues, expanding it for cyanobacteria.

In this article, we use a multi-lake data set of 822 Northern European lakes and evaluate trends in Cyanobacteria Biomass (CBB) using nutrient concentrations as predictors fitted with a non-parametric Generalised Additive Model (GAM) curve and a LOWESS curve. Then, by dividing lakes into 10 groups with different physico-chemical characteristics, we implement a linear Bayesian hierarchical modelling framework and obtain posterior probability simulations that exhibit a strong predictive modelling performance overall that varies depending on lake group and number of observations. Results are implemented for analyzing lake CBB concentrations according to the three risk levels associated to human health for recreational activities (Low—CBB $\leq$ 2mg/L; Medium—CBB between 2 and 10 mg/L and High—CBB $>$ 10mg/L), as defined by the World Health Organization (2006). Finally, exceedance probability response surfaces are produced for a range of nutrient concentrations for the WHO risk levels, showing that the Bayesian hierarchical modelling framework can be used for lake eutrophication management, by setting nutrient targets to sustain specific CBB thresholds with an associated exceedance risk level.

2. Materials and methods

2.1. Dataset

Our dataset consists of a range of biological (cyanobacteria biomass, Chl-a), physical (latitude, altitude, surface area, mean-max depth, mean-max air temperature) and chemical (total nitrogen, total phosphorus, total nitrogen to total phosphorus ratio, alkalinity type and humic type) features of several Northern European lakes, extracted from the central database of the EU-funded project WISER (Moe et al., 2013). WISER was launched in 2009 and for three years, 25 European Institutions representing 16 countries have addressed the assessment and management of rivers, lakes, transitional and coastal waters in Europe. Although the dataset originally contained observations for several features for 1851 lakes, it was unbalanced in terms of the number of monitored features per lake. Thus, after a thorough screening procedure we ended up with a subset of 822 lakes containing data for all the aforementioned variables. In other words, even though initially we had data for 1851 lakes for a large number of variables, we ended up with a compact dataset of 822 lakes containing data for a smaller group of variables—our goal was to create a dataset with features that would be covered by all lakes. The final subset contains lakes from six Northern European countries, namely UK, Denmark, Norway, Sweden, Finland and Lithuania. The total observations are 4,175 from May to October and from 1980 to 2009. However, observations are unevenly distributed among years, months and lakes. A total of 164 lakes have only a single observation, while the rest of the lakes range between 2 and 55 observations. Approximately 30% of the observations are from August while 27%, 18%, 13%, 9% and 3% are from July, June, May, September and October, respectively. In Fig. 1 the spatial distribution of all lakes in the dataset across the European map is shown. More details on the data set are included in Mellios et al. (2020).

2.2. Categorizing lakes into groups

Lake-type-specific models rely on the simple assumption that lakes belonging to a specific group are likely to exhibit similar behavior and response to changes in intra and extra-lake conditions. Under this context, the response of CBB to stressors is expected to follow a similar behavior among lakes of the same type. In this

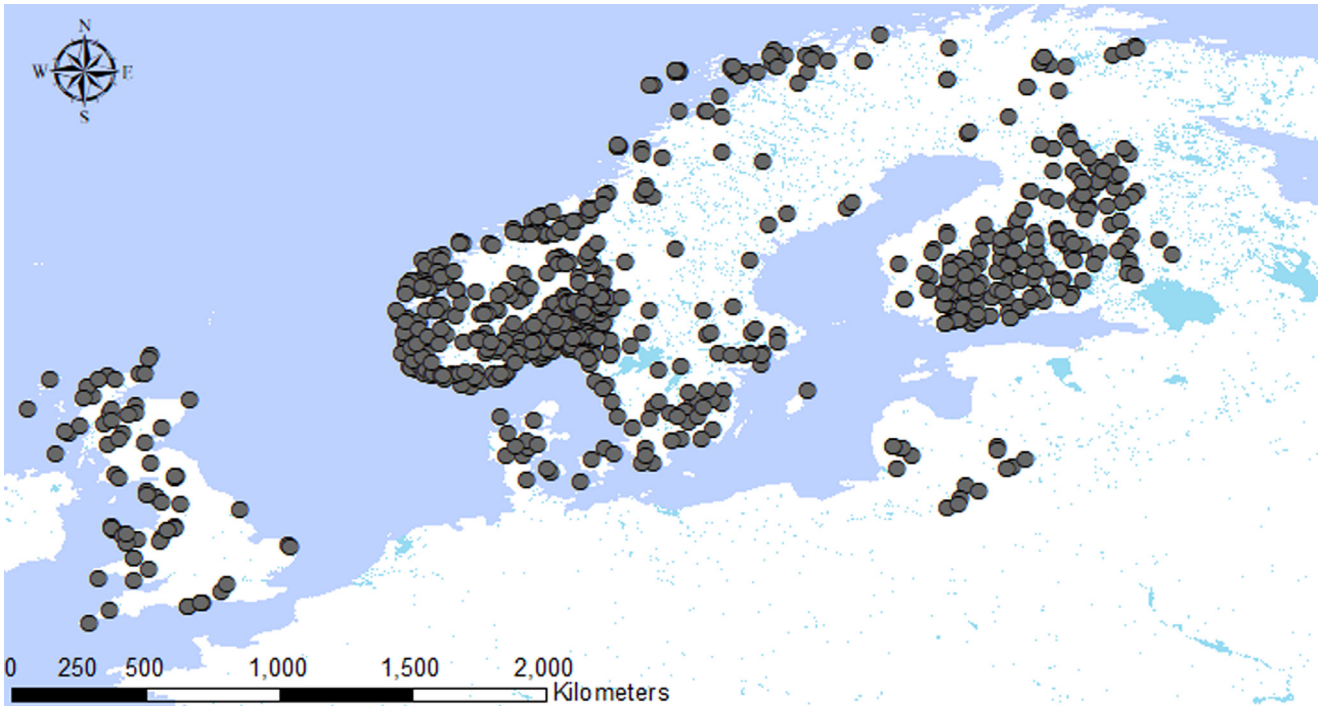


Fig. 1. Spatial distribution of lakes contained in the dataset.

work, grouping of lakes into types should ideally follow the lake typology defined for the Water Framework Directive (WFD) implementation within the Nordic Geographic Intercalibration Group (Poikane, 2009); Solheim et al. (2019) have developed a simplified version of the original typology—new broad typology—that has been used as guidance for developing lake groups in this article. Richardson et al. (2018) used a modified lake typology that included a classification in 18 lake types, which was eventually aggregated to 8 lake types to match data availability in each category. Malve and Qian (2006) used the Geomorphological Typology of Finnish Lakes specified by the Finnish Environment Institute, since their analysis included solely Finnish lakes. According to this typology, lakes are grouped into different types based on their geographical and natural characteristics (Pilke et al., 2002).

The WFD Nordic lake typology is strongly influenced by the Finnish lake typology, so our grouping was an adaptation of the latter to include criteria that match our dataset, which has a high proportion of Finnish lakes. Two typology variables, altitude and alkalinity or calcium level (siliceous vs. calcareous), are not included in the Finnish lake typology and were also left out from our analysis. The reason for excluding altitude was that only a few lakes had higher altitude, so this would result in the formation of unbalanced datasets under each lake category. Regarding Calcium level/alkalinity, such data were not available in our dataset for all lakes. Besides, alkalinity tends to co-vary with TP (although not for all lake groups) and its role has been investigated in other papers, namely in Richardson et al. (2018) and Carvalho et al. (2013). Here, we wanted to focus more on other factors, such as humic type, since there is evidence that there is a negative effect of humic level for cyanobacteria (Ptacnik et al., 2008; Richardson et al., 2018).

The chosen lake classification is very similar to the one used by Malve and Qian (2006) and included 10 groups, modifying lake types in order to fit the availability of data in our dataset. As specified in Table 1, the grouping of lakes was determined by mean depth, humic type and surface area. Mean depth was already discretized in the WISER dataset in three levels, namely “very shallow”, “shallow” and “deep”. In terms of humic type, the classes followed a similar discretization, namely “non- humic”, “humic” and “very humic”, indicated by the color level. In terms of lake size, the WISER typology included four classes according to lake surface area, namely “very small”, “small”, “medium” and “large”. In our dataset, we retained “large” and grouped “medium”, “small” and “very small” under a single category named “medium/small” and

Table 1
The adapted Geomorphological typology of lakes specified by the Finnish Environmental Agency. “D” refers to mean depth (m), “color” to humic type (mg Pt/L) and “SA” to surface area (km^2).

Lake Group	Explanation	Characteristics
1	very shallow, non-humic	D = 0 - 3 m, Color < 30
2	very shallow, humic	D = 0 - 3 m, Color > 30 and < 90
3	very shallow, very humic	D = 0 - 3 m, Color > 90
4	shallow, non-humic	D = 3 – 15 m, Color < 30
5	shallow, humic	D = 3 - 15 m, Color > 30 and < 90
6	shallow, very humic	D = 3 - 15 m, Color > 90
7	large, non-humic	SA > 10 km^2, Color < 30
8	large, humic	SA > 10 km^2, Color > 30 and < 90
9	medium/small, deep, non-humic	SA = 0 – 10 km^2, D > 15 m, Color < 30
10	medium/small, deep, humic	SA = 0–10 km^2, D > 15 m, Color > 30 and < 90

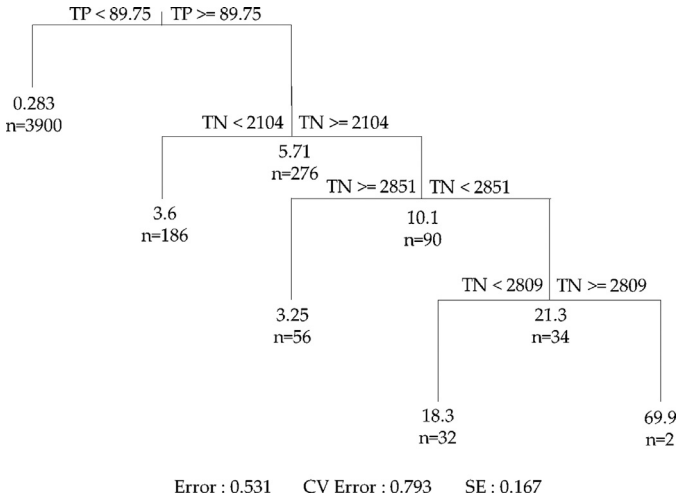


Fig. 2. CART analysis tree plot partitioned with TP and TN concentrations (μg/L).

ended up with two size categories. This grouping was done in order to reflect the types of lakes included in our dataset; a critical mass of data is ensured in each type with this grouping.

2.3. Identifying the best predictors of CBB

Linear correlation analysis indicated high correlation between CBB and Chl-a (r=0.52) (Mellios et al., 2020), but Chl-a was excluded from the explanatory variables because Chl-a and cyanobacteria are not independent, since cyanobacteria is a proportion of the total phytoplankton biomass, of which Chl-a is an indicator. Maximum air temperature and maximum depth show high collinearity with mean temperature and mean depth respectively; thus, they were also excluded since their effects on CBB are not distinct. To determine which of the other variables (latitude, altitude, surface area, mean depth, TN, TP, TN/TP and mean temperature) explain most of the variation of the response variable CBB, a classification and regression tree (CART) analysis was conducted. By satisfying the criterion to diminish the prediction error, the best tree model was chosen, which is illustrated in Fig. 2. The CART analysis procedure was done in the programming environment R version 3.6.2 (https://www.R-project.org/), by using the "*mvpart*" package (De'ath, 2007).

As indicated by the CART analysis results, when considering the whole dataset including all ten lake groups, TP plays the most significant role towards the prediction of CBB, while TN is influential only for the subset of samples (n = 276) where TP is larger than 89.75 μg/L. These findings are in accordance with Downing et al. (2001) who analyzed data from 99 temperate lakes and showed that cyanobacteria blooms are more strongly correlated with variation in TP and TN than the ratio of TN/TP. Watson et al. (1997) performed regression analyses and showed that the mean summer biomass of cyanobacteria is significantly and positively related to TP, for different nutrient ranges. Håkanson et al. (2007) used linear regression to model a transformed form of cyanobacteria concentration ($CBB^{0.25}$) on log(TN) and log(TP) and found significant correlations with R^2=0.63 and 0.75, respectively and concluded that variations in TP rather than TN seem to be more important to predict variations among systems in cyanobacteria. Following these results, TN and TP concentrations were selected as the best predictors of CBB and were used to construct the Bayesian hierarchical linear regression model. The advantage of including only two predictors is that the model is simple, lean and flexible and easy to be applied.

2.4. Bayesian hierarchical linear regression model

Bayesian methods are gaining ground in a wide range of scientific fields and especially in ecology, mainly due to their ability to produce probabilistic-oriented inferences, which in many cases outperform deterministic approaches. Since ecological modelling is characterized by high uncertainty due to the complex and many times unknown cause-effect relationships among variables, a probabilistic approach that yields distributions of possible outcomes, in essence, transforms uncertainty into probability thresholds. The advantage of Bayesian methods relies on their ability to combine prior knowledge about model parameters with evidence from data (Arhonditsis et al., 2006). They are well suited for analysis of multilevel models, showing: i) flexibility in specifying multilevel structures of parameters using priors, ii) ability to handle small samples and model misspecification (overparameterization of the likelihood can be resolved with well-chosen priors), iii) explicit handling of uncertainty and iv) intuitive and easy interpretation of results (credible interval versus confidence interval) (Grzenda, 2015). The hierarchical modelling approach implemented in this work is shown below:

$$y_{ijk} \sim N\left(X\beta_{ij},\ \tau^2\right) \tag{1}$$

$$X\beta_{ij} = \beta_{0,ij} + \beta_{1,ij} * TN_{ijk} + \beta_{2,ij} * TP_{ijk} \tag{2}$$

$$\beta_{ij} \sim N\left(\beta_i,\ \sigma_i^2\right) \tag{3}$$

$$\beta_i \sim N\left(\beta,\ \sigma^2\right) \tag{4}$$

$$\beta \sim N(0,\ 10000) \tag{5}$$

$$\sigma_i,\ \sigma \sim gamma(0.001,\ 0.001) \tag{6}$$

$$\tau = unif(0,\ 100) \tag{7}$$

where, y_{ijk} is the k^{th} observed CBB value from lake j in group i. X is the model matrix consisting of observed TN and TP values from lake j in group i, $\beta_{ij} = [\beta_{0,ij},\ \beta_{1,ij},\ \beta_{2,ij}]$ is the lake-specific linear regression model parameter vector which includes the intercept ($\beta_{0,ij}$) and the slopes for TN ($\beta_{1,ij}$) and TP ($\beta_{2,ij}$), τ^2 is the model error variance, $\beta_i = [\beta_{0,i},\ \beta_{1,i},\ \beta_{2,i}]$ is the vector of model parameter means for lake group i, $\sigma_i^2 = [\sigma_{0,i}^2,\ \sigma_{1,i}^2,\ \sigma_{2,i}^2]$ is the vector representing the variance of model parameters among lakes belonging to group i, while $\beta = [\beta_0,\ \beta_1,\ \beta_2]$ and $\sigma^2 = [\sigma_0^2,\ \sigma_1^2,\ \sigma_2^2]$ are the means and variance among groups, respectively.

The hierarchy of the specified model relies on the assumption that each lake's CBB values are modelled conditional on lake-specific model parameter values; the lake-specific model parameter values are modelled conditional on a common distribution representing all group-specific lakes, the lake group-specific parameter values are modelled conditional on a common parameter distribution representing all groups; the ensemble of groups is modelled conditional on a common distribution representing all lakes, while all lakes are in turn modelled conditional on representative hyperparameters for the whole population of lakes considered in our dataset. To be more specific, y_{ijk} is conditionally normally distributed on $X\beta_{ij}$ and τ^2, β_{ij} are conditionally normally distributed on β_i, σ_i^2, while β_i is conditionally normally distributed on β and σ^2. The non-informative prior distributions of β, τ, σ_i and σ which represent the hyperparameters follow a normal distribution $N(0,\ 10000)$ with mean 0 and variance 10,000, a uniform distribution unif(0, 100) with lower (0) and upper (100) limits and a gamma distribution gamma(0.001, 0.001) with shape parameter k (0.001) and scale parameter θ (0.001), respectively. The hyperparameters of τ, σ_i and σ are considered "vague" or non-informative as there is no information about their distribution.

2.5. Description of the modelling procedure

The Bayesian hierarchical analysis was conducted with the WinBUGS software (Lunn et al., 2000) which is a program for Bayesian analysis of complex statistical models using Markov Chain Monte Carlo (MCMC) techniques. In this study, the Metropolis algorithm was used which is based on a symmetric normal proposal distribution, whose standard deviation is turned over the first 4000 iterations in order to get an acceptance rate of between 20% and 40% (Lunn et al., 2000). To run the constructed model, a chain was produced and run for 100,000 iterations in order to let the MCMC simulation converge to the true posterior distribution. To check the convergence of the proposed model, we used the Heidelberger and Welch diagnostic, which is appropriate for the analysis of individual chains, under the "*BOA*" package in the programming environment R, version 3.6.2 (Smith, 2007). The advantage of this diagnostic method is two-fold; it both estimates the number of samples to be discarded as a burn-in sequence and it tests for non-convergence. In our case, the burn-in period as indicated by the convergence diagnostic tool was 50,000, while convergence was succeeded over 100,000 iterations. In order to reduce autocorrelation of the sample we took 1,250 samples for each unknown parameter (β_{ij}, β_i, σ_i^2, σ^2, τ^2) from the 50,000 remaining MCMC iterations by keeping the data of every 40th iteration (thin = 40). Finally, we confirmed the accuracy of the posterior parameter values by assuring that the MC error to the sample standard deviation error ratio for all parameters did not exceed the 5% limit, as proposed by Spiegelhalter et al. (2002).

3. Results and discussion

3.1. Data exploratory analysis

In Table 2 and Fig. 3, we show data statistics and boxplots of the response and explanatory variables included in the analysis for all lake groups. Exploring the relationship between CBB and nutrients in each group, it is noticeable that this relationship varies among groups. In groups 1 to 3, mean CBB decreases as humic level increases, which is consistent with the finding that cyanobacteria dominate more often in clear lakes than in humic ones (Ptacnik et al., 2008). Lake depth (groups 9 and 10) seems to play a determinant role in minimizing cyanobacteria abundance, which is in line with several studies (Bakker and Hilt, 2015; Sharma et al., 2011). However, no clear pattern can be detected between the relationship of CBB and nutrients, even though higher mean TN and TP values result in higher CBB values for the most part. This is probably related to the variable carrying capacity (maximum abundance) of lakes for cyanobacteria and to the nutrient that is limiting in each lake type; thus, even though phosphorus is often considered the limiting nutrient in lakes (Richardson et al., 2018), nitrogen can also play a key role (Beaulieu et al., 2013).

Table 3
Statistical analysis for the intercept for log(TN) and log(TP) GAM models.

Intercept	Estimate	Std. Error	t-value	Pr(>|t|)
CBB ~ s(TP)	0.64214	0.04682	13.71	<2e-16
CBB ~ s(TN)	0.64214	0.04854	13.23	<2e-16

Table 4
Approximate significance of smooth terms.

	edf	Ref.df	F	p-value	R^2 (adj)	Deviance explained
s(TP)	8.255	8.821	110.2	<2e-16	0.189	19%
s(TN)	8.14	8.796	70.05	<2e-16	0.128	13%

Some of the basic properties of the data for all lakes are shown in Fig. 4(a) and (b), where we show a scatter plot of the relationships of CBB vs. log(TP) and log(TN) for all lakes fitted with a non-parametric Generalised Additive Model (GAM) curve, along with the 5 and 95% Confidence Intervals. Statistical analyses for both curves are shown in Tables 3 and 4, respectively. In Table 3, the intercept estimate and standard errors are very similar for both TP and TN, signifying that if either variable were zero, the model would predict the same value of 0.6421 mg/L for CBB. In terms of the significance of the smooth terms (Table 4), again both variables have similar results, with TP performing better, with 19% of deviance explained as opposed to 13% for TN. This is consistent with our CART analysis that showed that TP plays the most significant role towards the prediction of CBB, while TN is influential only for a subset of samples. This result is also consistent with other research works, e.g. Hamilton et al. (2016), Søndergaard et al. (2017) and Moss et al. (2013). The p-values (Table 4) are very small in both cases for intercepts and smooth terms, showing that the data is sufficient to recognize that the smoothed relationship between CBB and TN and TP explains the data better than assuming that CBB is independent of TN and TP (intercept-only model).

To further explore the relationship between CBB and both TN and TP, we perform a LOWESS curve-fitting analysis. Here, CBB is predicted using two explanatory variables, i.e. CBB ~ s(TP, TN) and the 3D scatter plot with the predicted surface is shown in Fig. 5(a), while the corresponding contour plot is shown in Fig. 5(b). The model does not predict peak CBB concentrations for the combination of the highest TN and TP concentrations, but rather for relatively low TP concentrations (in the order of 200 μg/L) and medium/low TN concentrations (in the order of 2,300 μg/L). This signifies the fact that, depending on the concentrations, there is TN and TP limitation and lakes reach maximum abundance at those TN and TP concentrations. The model fails to capture most of the high CBB values, since peak CBB values in the order of 15 mg/L are predicted, while peak observed values are in the 70s. This is expected however, since model predictions are based on the

Table 2
Number of lakes, number of observations, mean (± standard deviation) of observed CBB, TN, and TP, within the lake groups.

Lake Group	Number of lakes	Obs.	Mean CBB (mg/L)	Mean TN (μg/L)	Mean TP (μg/L)
1	45	248	3.022 (±8.422)	915.519 (±662.166)	67.474 (±98.128)
2	74	340	2.360 (±6.652)	1401.567 (±1338.741)	95.345 (±129.138)
3	24	86	0.135 (±0.801)	636.035 (±189.824)	27.421 (±11.578)
4	208	1162	0.384 (±1.507)	688.353 (±769.757)	24.710 (±87.993)
5	126	768	0.429 (±2.716)	533.188 (±335.010)	20.012 (±26.187)
6	31	153	0.428 (±1.584)	594.749 (±330.709)	31.867 (±33.195)
7	97	340	0.082 (±0.337)	308.321 (±180.910)	8.686 (±8.998)
8	110	464	0.351 (±2.367)	548.891 (±343.360)	18.737 (±22.352)
9	91	515	0.152 (±0.835)	518.285 (±689.939)	14.249 (±31.071)
10	16	99	0.074 (±0.321)	578.359 (±416.881)	10.543 (±7.964)

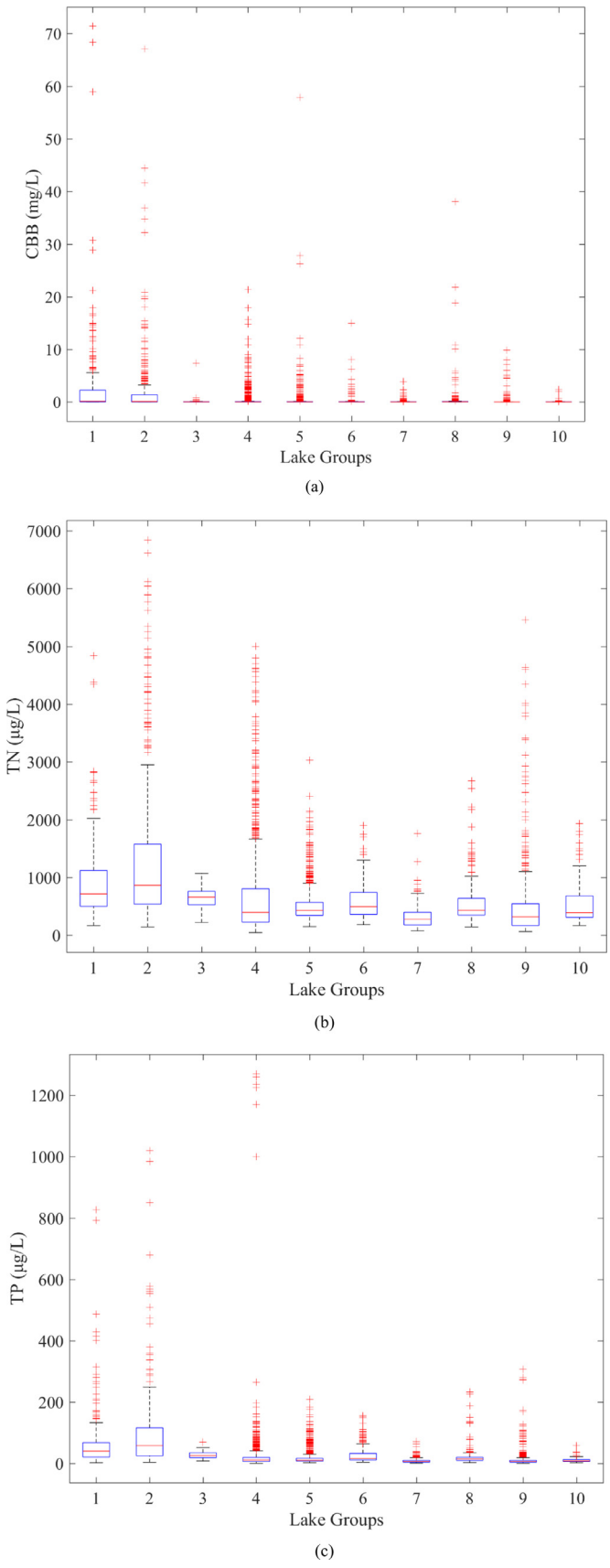


Fig. 3. Box-plots for all the lake groups: (a) CBB, (b) TN, and (c) TP.

whole CBB dataset—a skewed distribution with many zero values, as shown in Figure 3(a). Statistics of the analysis are shown in Fig. 5; a relatively low R^2 is obtained (similar to what has been observed in the literature, i.e Beaulieu et al., 2013; Carvalho et al., 2013; Chirico et al., 2020), but it is improved from the individual CBB vs. TN and CBB vs. TP plots (Table 4), proving that the capability of prediction of CBB using these advanced smoothing models is limited and predictions have relatively low reliability.

3.2. Linear hierarchical modelling and model fit

Having explored the potential to predict cyanobacteria with nonlinear relationships, we proceed with a linear hierarchical model, to take advantage of the simplicity and flexibility of linear models. Linear regression is used, even though cyanobacteria response to the nutrient gradient is non-linear (Fig. 4). Nevertheless, linear models are still commonly used in ecology, since nonlinear models tend to become complicated and require the predefinition of parameters by the user, such as defining the maximum of the curve, or the point where the concavity and/or convexity of the curve begins (Carvalho et al., 2013); this way, there exists a potential to introduce error by predefining the results. With skewed distributions, the assumptions underlying regression models based on normal distribution are violated, so data transformation is commonly used. The Box-Cox procedure defines the parameter λ that is used to choose the most suitable transformation for the dataset to achieve normality. For our dataset, λ was close to zero, so the suitable transformation would be logarithmic. This is relatively common in ecology, since environmental variables take only positive values and the logarithm of these variables are likely to be normal and the resulting model is easy to interpret (Qian, 2016). In this work, a log transformation did not prove useful, first because the dataset is unbalanced and includes a large number of zero concentrations of CBB. Even when replacing zero concentrations with small numbers to avoid the conflict with the log transformation, the retransformation of the log CBB variable was problematic, because the exponential of the log-mean was not the same as the mean concentration and model fit deteriorated, when compared to the untransformed dataset. In this case, the error term ε cannot be ignored and needs to be included in the retransformation of data, ultimately introducing a bias to the results by the fixed multiplicative factor e^{ε}. Even though using a bias correction factor is a possibility (Sprugel, 1983), it is difficult to define a formula for the standard error and the estimated mean of the dependent variable (Qian, 2016). Based on this analysis, we conducted the modelling with untransformed data and obtained good model fits overall with hierarchical modelling. Log-transforming TN and TP in order to setup a linear-log model did not improve results; therefore, all variables were used untransformed.

To assess hierarchical modelling fit in terms of both precision and accuracy, we compare predicted vs. observed CBB values through scatter plots for all lake groups, as shown in Fig. 6. The circles represent the mean predicted values while the lower and upper limits of the red lines are the 10^{th} and 90^{th} percentiles. In the last graph, we show simulated vs. observed for the whole dataset (all lakes). At the value of 0.74, we see that R^2 for all lakes (Fig. 6) is remarkably improved when compared to the predictions that were reported for multiple linear regression for the same data set in Mellios et al. (2020) (R^2=0.33). Predictions are also greatly improved when compared to both non-parametric curve models performed (GAMS or LOWESS). Lake groups show variable accuracy and precision, with some groups performing impressively well (e.g., R^2=0.87 for Group 8—large humic lakes; R^2=0.85 for Group 6—shallow very humic lakes and R^2=0.81 for Group 1—very shallow, non-humic lakes), while some groups performing poorly (e.g., R^2=0.15 for Group 10—medium/small, deep, humic lakes). It seems

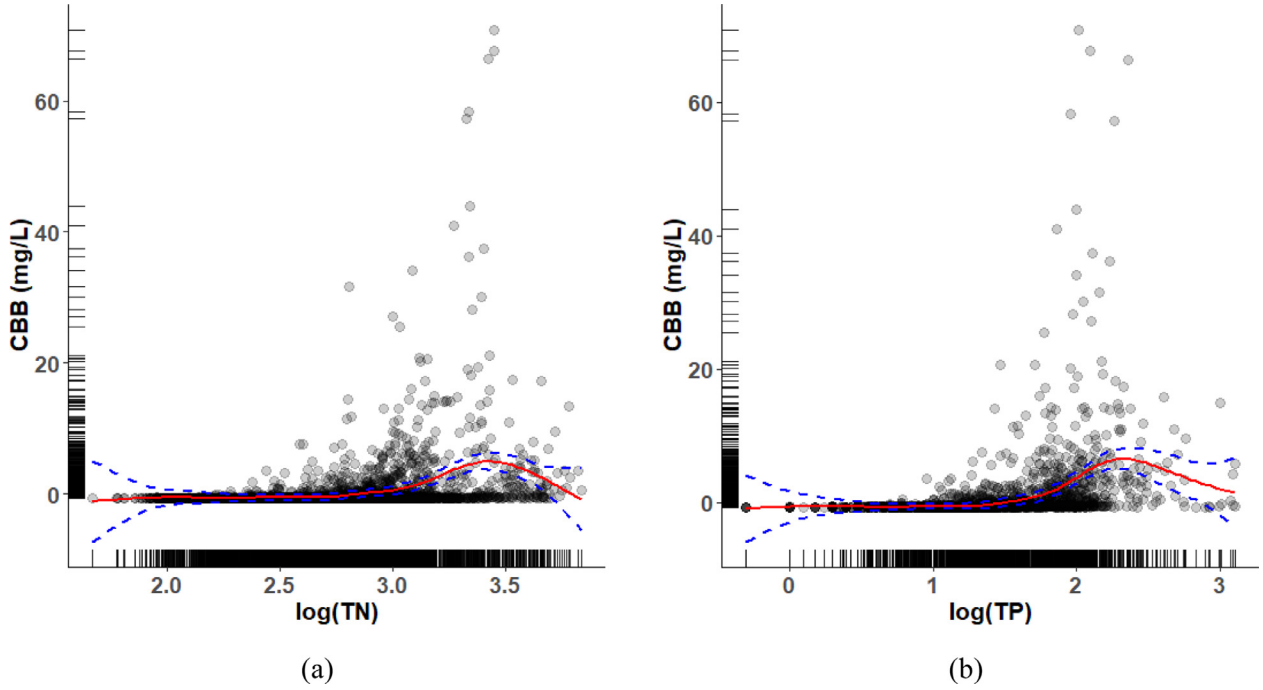
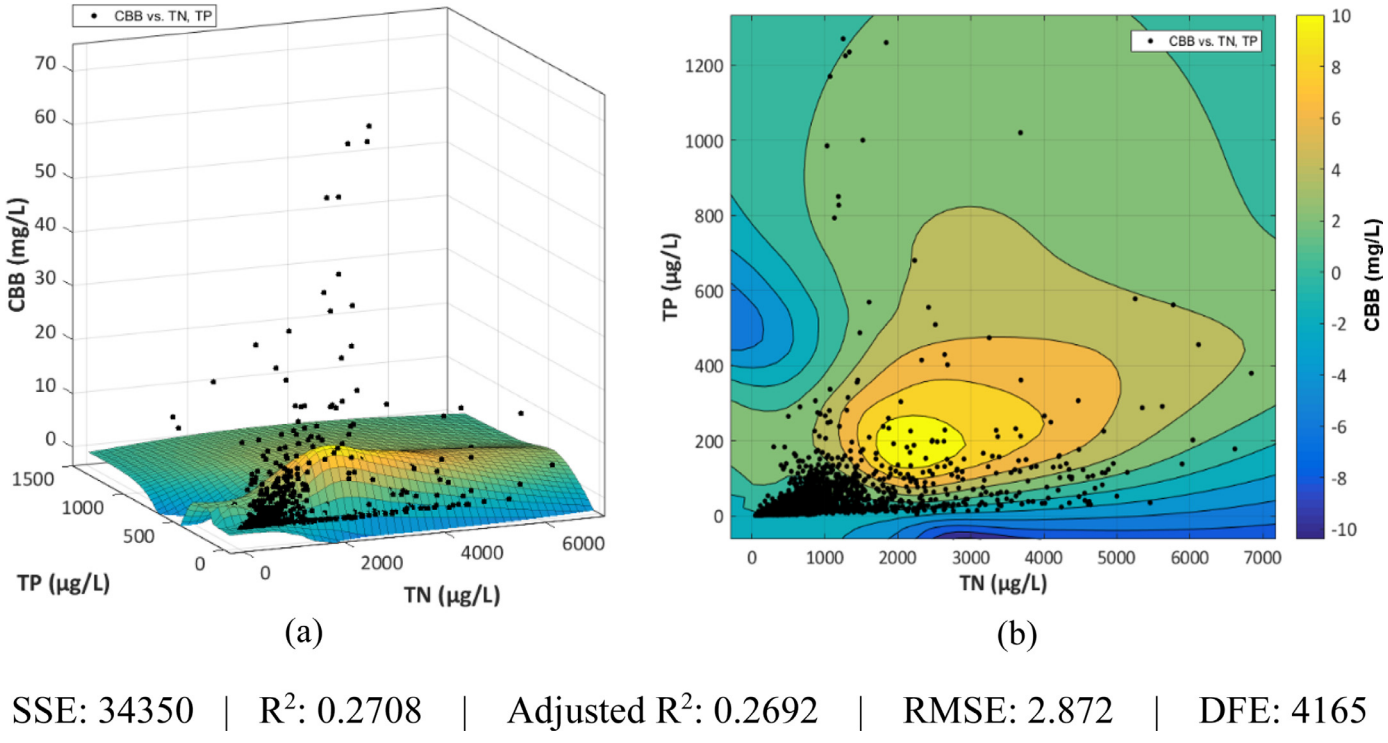


Fig. 4. Scatter plot of (a) CBB vs log(TN) and (b) CBB vs log(TP) with smoothing curves and 5% and 95% Confidence Interval curves using GAM.



SSE: 34350 | R^2: 0.2708 | Adjusted R^2: 0.2692 | RMSE: 2.872 | DFE: 4165

Fig. 5. CBB vs. TN and TP with LOWESS analysis: (a) 3D scatter plot and predicted surface and (b) associated contour plot.

that the model drops in performance when the maximum observed CBB values in a group are small, i.e., less than 3 mg/L, as is the case in group 10; the number of observations is also important with the number of observations being inversely proportional to model fit.

The hierarchical model, even though it models the non-linear response of CBB to nutrient concentrations linearly, has a great advantage in the fact that while it fits different model parameters for each lake, thus customizing the model to the specific lake data, it also treats lakes within the same group as exchangeable. Essentially, parameters of lakes in the same lake group are assumed to come from the same prior distributions, thereby pooling information from similar lakes. As a result, this pooling of information results in reduced bias at lake-level, while model error variance is reduced as well. This method is superior especially for lakes that need to be managed for eutrophication but have no prior data

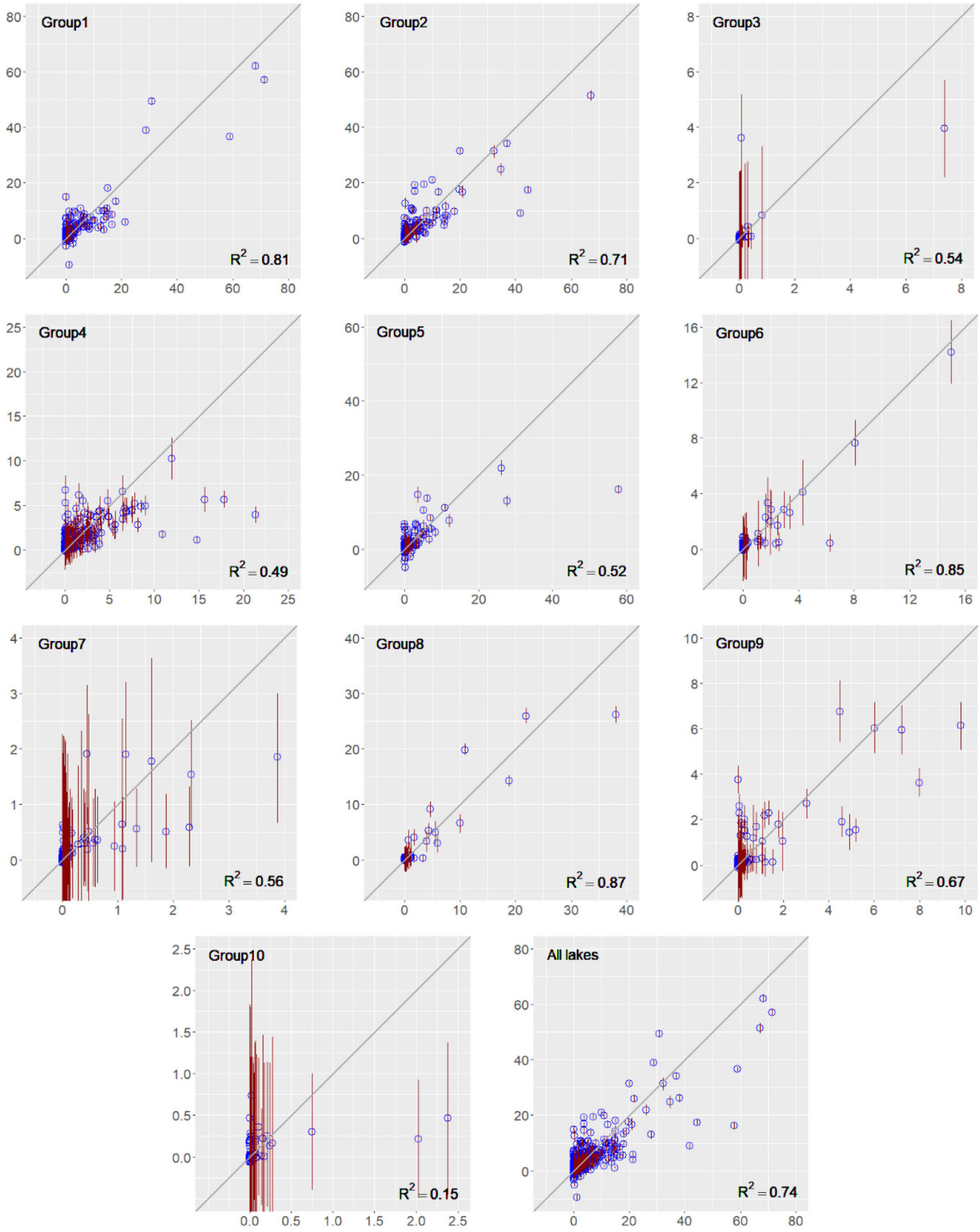


Fig. 6. Scatter plots for mean predicted vs. observed CBB values (in mg/L) for the 10 lake groups and the whole set of lakes, produced by the linear Bayesian hierarchical model. Predicted values are shown on the y-axis and observed values on the x-axis.

to base management decisions on: The lake will be classified in a group according to its characteristics and the group parameters will be used to model it. As more data are included, the lake model will improve by customising its model parameters and differentiating them from those of the group. However, during the phase that no data exist, there will be an initial set of parameters to be used successfully.

3.3. Posterior probabilities and exceedance probability surfaces

In order to demonstrate how hierarchical models can be used for the management of eutrophic lakes, we simulate posterior probabilities of CBB for a range of TN and TP for an indicative list of four specific lakes coming from groups 1, 6, 8 and 10 (Engelsholm Sø, Lillsjön, Päijänne and Gjersjøen, respectively). The number of observations for these lakes are 18, 3, 24 and 11, respectively. The TN and TP ranges that are plotted are consistent with the corresponding observed concentrations for each lake. In Figs 7(a), (d), (g) and (j), the resulting surface is presented for the 50th percentile of the predictive distributions along with the scatter plots of observed values in a 3D format, showing the goodness of fit of the model. These surfaces were designed using the posterior distribution parameters β_{ij} ($\beta_{0,ij}$, $\beta_{1,ij}$, $\beta_{2,ij}$) that are specific for each lake. In Figs 7(b), (e), (h) and (k), we show the same plots as before, but without the observed values scatter plot; this time we visualize two horizontal planes that correspond to the three distinct health risk levels (low—medium—high, with thresholds at 2 and 10 mg/L), as defined by WHO, after converting cell counts into concentrations (Mellios et al., 2020). From these posterior probability surfaces, we can identify the combination of TN and TP concentrations that result in the 50th percentile of CBB distribution being lower than 2 mg/L (below the bottom horizontal plane—green color), being between 2 and 10 mg/L (between the two planes—yellow color) and being above 10 mg/L (above the top horizontal plane—red color). Posterior probability surfaces are shown for all four lakes, showing how successful hierarchical modelling is in capturing the variability of CBB for a wide range of lakes and lake groups and multitude of observations. In Figs 7(h) and (k), the predicted concentrations are below the 2 mg/L threshold; thus, no horizontal plane is shown. It should be noted here that even though for all groups, plots in the left and center panels are identical, for groups 1, 6 and 10, the plots in the center panel are rotated for better visualization of the 3-D surfaces. In the right panels—Figs 7(c), (f), (i) and (l)—we show the 3D posterior probability surface plots for the 50th percentile of predictive distributions for each lake group, for TN and TP ranges consistent with the corresponding observed concentrations for all lakes belonging to each group. This way, we can see how hierarchical modelling works: In the absence of any data for a single lake, the prediction probabilities would have to come from the lake group level in which the lake belongs—shown in Figs 7(c), (f), (i) and (l). These surfaces were designed using posterior distribution parameters β_i ($\beta_{0,i}$, $\beta_{1,i}$, $\beta_{2,i}$) for each lake group. Naturally, as data from the lake become available, hierarchical modelling greatly improves and provides more accurate predictions specific for each lake, as shown in the center and left panels.

Table 5 lists all posterior distribution parameters β_i ($\beta_{0,i}$, $\beta_{1,i}$, $\beta_{2,i}$) for all ten lake groups. When comparing $\beta_{1,i}$ (the coefficient of TN) and $\beta_{2,i}$ (the coefficient of TP), we see that for almost all lake groups, $\beta_{2,i}$ is positive and over one order of magnitude greater than $\beta_{1,i}$. Indeed, this result matches with the CART presented earlier, since it enforces that TP is the most influential, being the most important one strongly affecting CBB; TN is also important but seems to play a secondary role.

The Bayesian hierarchical model becomes a powerful framework when CBB standard exceedance probability response surfaces are simulated. Such a surface can set nutrient criteria that can be directly used under a risk assessment framework for eutrophic lake management. Therefore, an exceedance probability response for the 10 mg/L CBB threshold shows the range and combination of TN and TP concentrations for which the predicted CBB has less than a specific probability to exceed that threshold of 10 mg/L; this is indicatively shown in Fig. 8(a) for lake Engelsholm Sø (Group 1), where x and y axes show nutrient concentrations and the z-axis is the probability to exceed the 10mg/L threshold. In Fig. 8(b) and 8(c), we show the contour diagram of the exceedance probability response surface for all percentiles, for the 2 mg/L and 10 mg/L threshold, respectively. For a lake eutrophication management scheme, the lake manager can identify the risk level that (s)he wants to operate under. If a 90% risk level is chosen, then the combination of TN and TP concentrations that correspond to the 90% line in Fig. 8(b) and (c) signify the concentrations that give a 90% probability to exceed the 2 mg/L and 10 mg/L thresholds, respectively. For a lower risk level, lower TN and TP concentrations are required. Alternatively, if TN and TP concentrations in the lake are measured under a monitoring scheme, or if criteria for TN and TP are set by the European Water Framework Directive (Poikane et al., 2019) or a relevant authority, the lake manager can have an estimate of what the risk level is to exceed the 2 or 10 mg/L threshold. To show how this might work, in Fig. 8(b) and (c), we plot the observed combinations of TN and TP concentrations for lake Engelsholm Sø along with the corresponding CBB concentrations. With red font, we show the CBB concentrations that exceed the preset thresholds and we see that indeed all high CBB concentrations appear in the area that is above the 97.5% exceedance line. Only a few high CBBs are found on the risk level lines 25% and higher, while there is no “red font” in the “safe” area under the 2.5% risk line.

With posterior predictive simulations of the Bayesian hierarchical approach, similar curves can be drawn for any threshold, providing a flexible and robust framework of probabilistic risk assessment for various management decisions and associated nutrient concentrations that is relatively easy to use and understand. To our knowledge, these contour diagrams of the exceedance probability response surface for a given threshold of CBB for all percentiles and for the full range of nutrient concentrations in a single graph as a result of Bayesian hierarchical modelling has not been done before and is a novel and powerful methodology for lake eutrophication management. The Lake Load Response model, also a Bayesian hierarchical model, was developed into an online tool for lake managers (http://lakestate.vyh.fi) and can be used e.g. for prediction of Chl-a and phytoplankton biomass from TN and TP loads. LLR is a useful lake management tool that allows the calculation of estimates of the amount of loading reduction needed to achieve good water quality in a lake. In a similar way, our model can form the basis of a tool for predicting cyanobacteria biomass from TN and TP concentrations for different lake groups.

Table 5
The 50th percentile posterior distribution parameters for the 10 lake groups.

Lake Group	β_0	β_1	β_2
1	-0.013400	0.003378	0.012380
2	-0.022910	0.000209	0.022720
3	0.000905	0.000305	0.002986
4	-0.044540	0.000344	0.011540
5	-0.011460	0.000605	0.003927
6	-0.008862	-0.000054	0.014920
7	-0.005495	-0.000001	0.006400
8	-0.009015	0.000888	-0.012680
9	-0.018710	-0.000043	0.009474
10	0.000656	-0.000187	0.005845

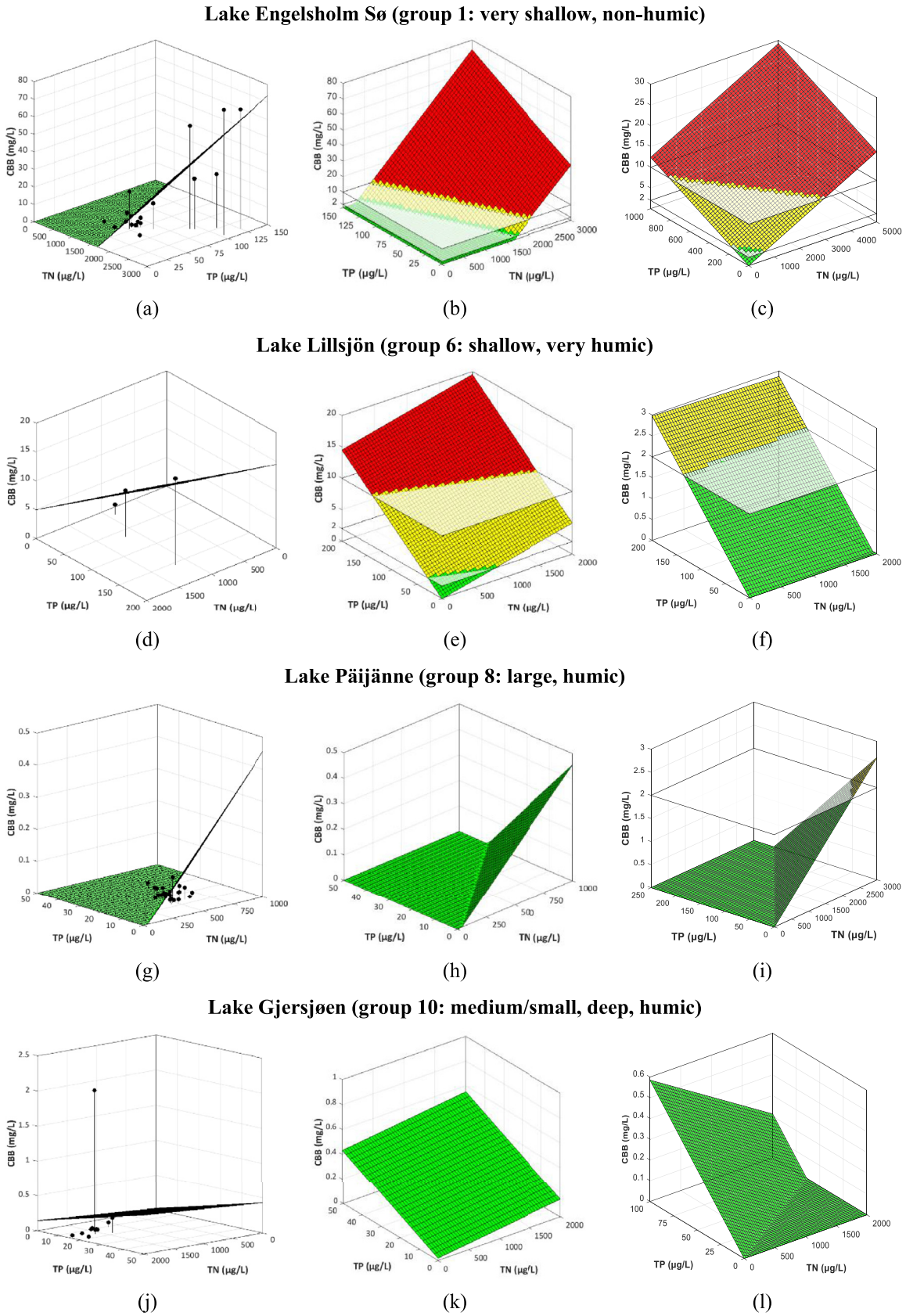
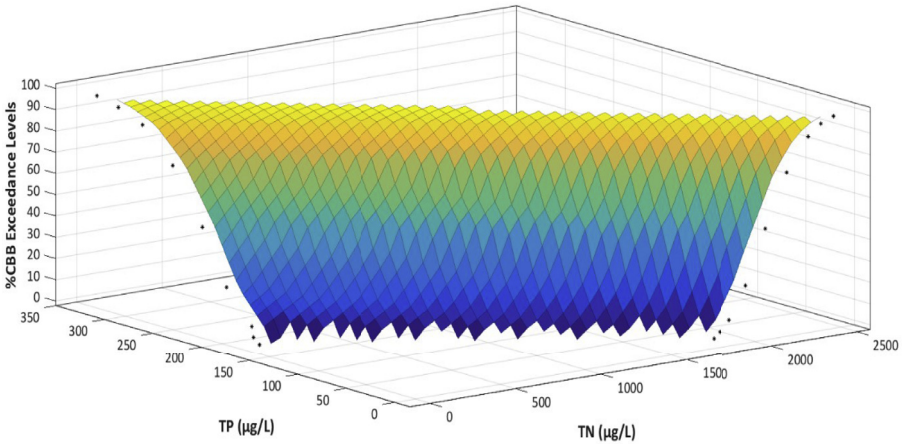
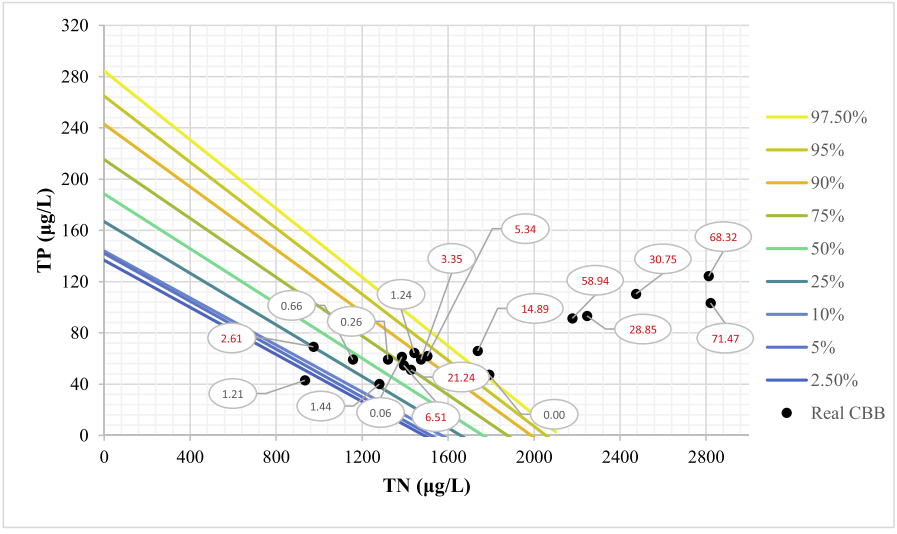


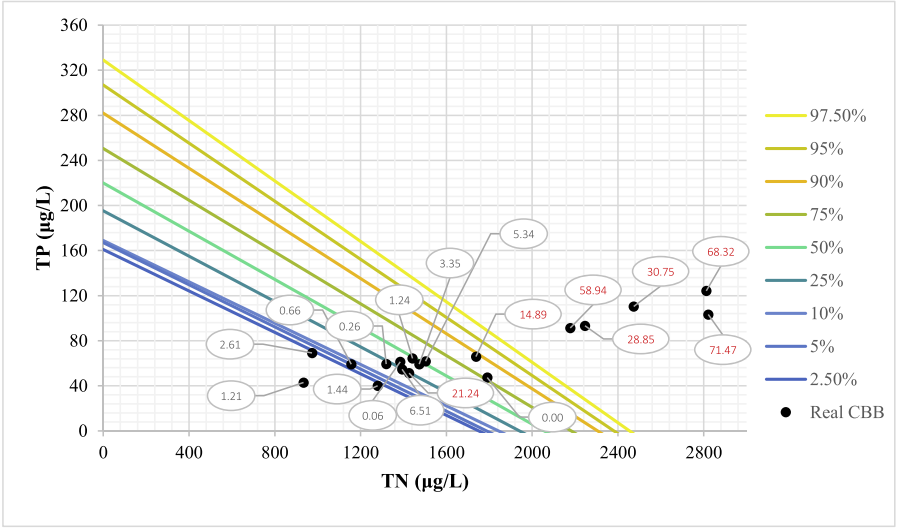
Fig. 7. 3D posterior probability surface plots for the 50th percentile of the predictive distributions along with scatter plots of observed values (left panels) for four different lakes, produced by the linear Bayesian hierarchical model. In the center panel of each row, the same surface is shown along with two horizontal planes corresponding to the two thresholds that define low-medium and high-risk levels. In the right panel for each row, the 50th percentile of the predictive distributions are shown for each corresponding lake group, for comparison purposes. It should be noted that to achieve best visualization, surface plots in the middle panel of lake groups 1, 6 and 10 have been rotated in order to provide a different perspective and facilitate visualization of observed values (points).



(a)



(b)



(c)

Fig. 8. **(a)** Exceedance probability response surface for the 10 mg/L CBB threshold versus TN and TP for Lake Enghelsholm Sø; **(b)** corresponding contour plot showing TN and TP concentrations and associated risk of exceedance for CBB concentrations of 2 mg/L and **(c)** 10 mg/L. Observed TN-TP concentrations are plotted along with their corresponding CBB concentrations. CBB concentrations in red actually exceed preset thresholds of (b) 2 mg/L and (c) 10 mg/L.

4. Conclusions

The Bayesian hierarchical linear regression model calculates lake-specific probabilities of CBB concentrations to exceed the two health risk levels for recreational use, under different TP and TN concentrations. Enabling lake managers to define combinations of TP and TN concentrations that will result in exceedance risk levels for pre-defined thresholds appropriate for each ecosystem can lead to optimal monitoring schemes and can minimize uncertainty associated with each lake ecosystem. After compiling a large water quality data set for a system of lakes with different characteristics and typology divided into groups, competent authorities can develop a monitoring strategy that will focus on lakes that have the greatest risk at violating the thresholds and developing cyanobacterial blooms. Thus, monitoring schemes become targeted and efficient with maximum benefits for society. Using Bayesian hierarchical modelling, lake managers can focus on lakes with significant ecosystem services, including recreational quality and can thus maximise the provision of these services. Categorizing lakes in different types according to their characteristics allows a clear generalization of lake responses, promotes our understanding of lake cyanobacteria dynamics and enables lake managers to target measures to minimize risks under climate change. Finally, a lake with no data history can take advantage of the data series of other lakes that belong to the same group and can follow a management scheme that will be superior to a generalized scheme that would be applicable to all lakes.

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

The work described inthis paper has been conducted within the project WATER4CITIES—Holistic Surface Water and Groundwater Management for Sustainable Cities—which is implemented in the framework of the EU Horizon2020 Program, Grant Agreement Number 734409. This paper and the content included in it do not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content. We thank the following organisations and persons for making phytoplankton and environmental data available through the WISER Central Database. UK: Environment Agency England & Wales (EA) and Scottish Environment Protection Agency (SEPA) (Geoff Phillips, Lawrence Carvalho); Norway: Norwegian Institute for Water Research (NIVA) (Anne Lyche Solheim, Birger Skjelbred); Sweden: Swedish University of Agricultural Sciences (SLU) (Stina Drakare); Finland: Finnish Environment Institute (SYKE) (Marko Järvinen); Denmark: Aarhus University (Martin Søndergaard, Ivan Karottki); Lithuania: Environmental Protection Agency Lithuania (AAA) (Audrone Pumputyte).

References

Arhonditsis, G.B., Stow, C.A., Steinberg, L.J., Kenney, M.A., Lathrop, R.C., McBride, S.J., Reckhow, K.H., 2006. Exploring ecological patterns with structural equation modelling and Bayesian analysis. Ecol. Modell. 19 (3–4), 385–409. doi:10.1016/j.ecolmodel.2005.07.028.

Bakker, E.S., Hilt, S., 2015. Impact of water-level fluctuations on cyanobacterial blooms: options for management. Aquat. Ecol. 50 (3), 485–498. doi:10.1007/s10452-015-9556-x.

Beaulieu, M., Pick, F., Gregory-Eaves, I., 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. 58 (5), 1736–1746. doi:10.4319/lo.2013.58.5.1736.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the water framework directive. Ecol. Indic. 18, 31–41. doi:10.1016/j.ecolind.2011.10.009.

Carvalho, L., McDonald, C., de Hoyos, C., Mischke, U., Phillips, G., Borics, G., Poikane, S., Skjelbred, B., Solheim, A.L., Van Wichelen, J., Cardoso, A.C., 2013. Sustaining recreational quality of European lakes: minimizing the health risks from algal blooms through phosphorus control. J. Appl. Ecol. 50 (2), 315–323. doi:10.1111/1365-2664.12059.

Charmichael, W.W., Boyer, G.L., 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. Harmful algae 54, 194–212. doi:10.1016/j.hal.2016.02.002.

Cheng, V., Arhonditsis, G.B., Brett, M.T., 2009. A revaluation of lake-phosphorus models using a Bayesian hierarchical framework. Ecol. Res. 25 (1), 59–76. doi:10.1007/s11284-009-0630-5.

Chirico, N., António, D.C., Pozzoli, L., Marinov, D., Malagó, A., Sanseverino, I., Beghi, A., Genoni, P., Dobricic, S., Lettieri, T., 2020. Cyanobacterial blooms in lake varese: analysis and characterization over ten years of observations. Water 12 (3), 675. doi:10.3390/w12030675.

De'ath, G., 2007. The mvpart package. http://cran.r-project.org/web/packages/mvpart.

Downing, J.A., Watson, S.B., McCauley, E., 2001. Predicting cyanobacteria dominance in lakes. Can. J. Fish. Aquat.Sci. 58 (10), 1905–1908. doi:10.1139/f01-143.

Grzenda, W., 2015. The advantages of Bayesian methods over classical methods in the context of credible intervals. Information Systems in Management 4 (1), 53–63.

Håkanson, L., Bryhn, A.C., Hytteborn, J.K., 2007. On the issue of limiting nutrient and predictions of cyanobacteria in aquatic systems. Sci. Total Environ. 379 (1), 89–108. doi:10.1016/j.scitotenv.2007.03.009.

Hamilton, D.P., Salmaso, N., Paerl, H.W., 2016. Mitigating harmful cyanobacterial blooms: strategies for control of nitrogen and phosphorus loads. Aquat Ecol 50, 351–366.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliot, M., Feld, C.K., Heiskanen, A-S, Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., van de Bund, W., 2010. The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Sci. Total Environ. 408 (19), 4007–4019. doi:10.1016/j.scitotenv.2010.05.031.

Ibelings, B.W., Fastner, J., Bormans, M., Visser, P.M., 2016. Cyanobacterial blooms. Ecology, prevention, mitigation and control: Editorial to a CYANOCOST special issue. Aquat. Ecol. 50 (3), 327–331. doi:10.1007/s10452-016-9595-y.

Jewett, E.B., Lopez, C.B., Dortch, Q., Etheridge, S.M., Backer, L.C., 2008. Harmful algal bloom management and response: assessment and plan. interagency working group on harmful algal blooms, hypoxia, and human health of the joint subcommittee on ocean science and technology. Washington, DC.

Lévesque, B., Gervais, M.C., Chevalier, P., Gauvin, D., Anassour-Laouan-Sidi, E., Gingras, S., Fortin, N., Brisson, G., Greer, C., Bird, D., 2014. Prospective study of acute health effects in relation to exposure to cyanobacteria. Sci. Total Environ. 466, 397–403. doi:10.1016/j.scitotenv.2013.07.045.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Comput. 10 (4), 325–337. doi:10.1023/A:1008929526011.

Malve, O., Qian, S.S., 2006. Estimating nutrients and chlorophyll a relationships in Finnish lakes. Environ. Sci. Technol. 40 (24), 7848–7853. doi:10.1021/es061359b.

Mellios, N., Moe, S.J., Laspidou, C., 2020. Machine learning approaches for predicting health risk of cyanobacterial blooms in northern european lakes. Water 12 (4), 1191. doi:10.3390/w12041191.

Moe, S.J., Schmidt-Kloiber, A., Dudley, B.J., Hering, D., 2013. The WISER way of organising ecological data from European rivers, lakes, transitional and coastal waters. Hydrobiologia 704 (1), 11–28. doi:10.1007/s10750-012-1337-0.

Moss, B., Jeppesen, E., Søndergaard, M., Lauridsen, T.L., Liu, Z., 2013. Nitrogen, macrophytes, shallow lakes and nutrient limitation: resolution of a current controversy? Hydrobiologia 710 (1), 3–21.

Obenour, D.R., Gronewold, A.D., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resour. Res. 50 (10), 7847–7860. doi:10.1002/2014WR015616.

Pilke, A., Heinonen, P., Karttunen, K., Koskenniemi, E., Lepistö, L., Pietiläinen, O.P., Rissanen, J., Vuoristo, H., 2002. Finnish draft for typology of lakes and rivers. In: Ruoppa, M., Karttunen, K. (Eds.), In Typology and Ecological Classification of Lakes and Rivers. Nordic Council of Ministers, TemaNord, pp. 42–43 2002.

Poikane, S. (Ed.), Water Framework Directive intercalibration technical report. Part 2: Lakes. EUR 23838 EN/2, European Commission Joint Research Centre, 2009. http://publications.jrc.ec.europa.eu/repository/handle/JRC51340, DOI: 10.2788/23415.

Poikane, S., Kelly, M.G., Herrero, F.S., Pitt, J.A., Jarvie, H.P., Claussen, U., Leujak, W., Solheim, A.L., Teixeira, H., Phillips, G., 2019. Nutrient criteria for surface waters under the European Water Framework Directive: Current state-of-the-art, challenges and future outlook. Sci. Total Environ. 695, 133888. doi:10.1016/j.scitotenv.2019.133888.

Ptacnik, R., Solimini, A.G., Andersen, T., Tamminen, T., Brettum, P., Lepistö, L., Willén, E., Rekolainen, S., 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. Proc. Natl. Acad. Sci. 105 (13), 5134–5138. doi:10.1073/pnas.0708328105.

Qian, S.S., 2016. Environmental and Ecological Statistics with R, 2nd Edition. CRC Press, Boca Raton, FL.

Richardson, J., Feuchtmayr, H., Miller, C., Hunter, P.D., Maberly, S.C., Carvalho, L., 2019. Response of cyanobacteria and phytoplankton abundance to warming, extreme rainfall events and nutrient enrichment. Global Change Biology 25 (10), 3365–3380. doi:10.1111/gcb.14701.

Richardson, J., Miller, C., Maberly, S.C., Taylor, P., Globevnik, L., Hunter, P., Jeppesen, E., Mischke, U., Moe, S.J., Pasztaleniec, A., Søndergaard, M., Carvalho, L., 2018. Effects of multiple stressors on cyanobacteria abundance vary with lake type. Global Change Biol. 24 (11), 5044–5055. doi:10.1111/gcb.14396.

Sharma, N.K., Tiwari, S.P., Tripathi, K., Rai, A.K., 2011. Sustainability and cyanobacteria (blue-green algae): facts and challenges. J. Appl. Phycol. 23 (6), 1059–1081. doi:10.1007/s10811-010-9626-3.

Shimoda, Y., Arhonditsis, G.B., 2015. Integrating hierarchical Bayes with phosphorus loading modelling. Ecol. Informatics 29, 77–91. doi:10.1016/j.ecoinf.2015.07.005.

Shimoda, Y., Watson, S.B., Palmer, M.E., Koops, M.A., Mugalingam, S., Morley, A., Arhonditsis, G.B., 2016. Delineation of the role of nutrient variability and dreissenids (Mollusca, Bivalvia) on phytoplankton dynamics in the Bay of Quinte, Ontario, Canada. Harmful algae 55, 121–136. doi:10.1016/j.hal.2016.02.005.

Smith, B.J., 2007. boa: an R package for MCMC output convergence assessment and posterior inference. J. statistical software 21 (11), 1–37. http://hdl.handle.net/10.18637/jss.v021.i11.

Solheim, A.L., Globevnik, L., Austnes, K., Kristensen, P., Moe, S.J., Persson, J., Phillips, G., Poikane, S., van de Bund, W., Birk, S., 2019. A new broad typology for rivers and lakes in Europe: development and application for large-scale environmental assessments. Sci. Total Environ. 697, 134043. doi:10.1016/j.scitotenv.2019.134043.

Søndergaard, M., Lauridsen, T.L., Johansson, L.S., et al., 2017. Nitrogen or phosphorus limitation in lakes and its impact on phytoplankton biomass and submerged macrophyte cover. Hydrobiologia 795, 35–48.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. Royal Statistical Soc.: Series b (statistical methodology) 64 (4), 583–639. doi:10.1111/1467-9868.00353.

Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64 (1), 209–210. doi:10.2307/1937343.

Stow, C.A., Cha, Y.K., Qian, S.S., 2014. A Bayesian hierarchical model to guide development and evaluation of substance objectives under the 2012 Great Lakes Water Quality Agreement. J. Great Lakes Res. 40, 49–55. doi:10.1016/j.jglr.2014.07.005.

Tromas, N., Fortin, N., Bedrani, L., Terrat, Y., Cardoso, P., Bird, D., Greer, C.W., Shapiro, B.J., 2017. Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. The ISME J. 11 (8), 1746–1763. doi:10.1038/ismej.2017.58.

UNEP, 2016. A Snapshot of the World's Water Quality: Towards a global assessment. United Nations Environment Programme, Nairobi, Kenya. 162 pp.

UN, 2019. The Sustainable Development Goals Report. UN, New York. 10.18356/55eb9109-en.

Watson, S.B., McCauley, E., Downing, J.A., 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol. Oceanogr. 42 (3), 487–495. doi:10.4319/lo.1997.42.3.048.

, 2006. In: Chapter 8. In Guidelines for safe recreational waters: Coastal and fresh waters, 1. WHO Publishing, Geneva, Switzerland, pp. 136–158 2003.