Application of Bayesian network including Microcystis morphospecies for microcystin risk assessment in three cyanobacterial bloom-plagued lakes, China

Article

Accepted Version

Shan, K., Shang, M., Zhou, B., Li, L., Wang, X., Yang, H. ORCID: https://orcid.org/0000-0001-9940-8273 and Song, L. (2019) Application of Bayesian network including Microcystis morphospecies for microcystin risk assessment in three cyanobacterial bloom-plagued lakes, China. Harmful Algae, 83. pp. 14-24. ISSN 1878-1470 doi: 10.1016/j.hal.2019.01.005 Available at https://centaur.reading.ac.uk/81862/

Published version at: https://www.sciencedirect.com/science/article/pii/S1568988319300113
To link to this article DOI: http://dx.doi.org/10.1016/j.hal.2019.01.005

Publisher: Elsevier

Application of Bayesian network including *Microcystis* morphospecies for microcystin risk assessment in three cyanobacterial bloom-plagued lakes, China

Kun Shan[a*], Mingsheng Shang[a,c], Botian Zhou[a,c], Lin Li[b], Xiaoxiao Wang[c,e], Hong Yang[d], Lirong Song[b,e*]

[a] *Big Data Mining and Applications Center, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Science, Chongqing 400714, China*

[b] *State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*

c CAS Key Lab on Reservoir Environment, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China

[d] *Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB, UK*

e University of Chinese Academy of Sciences, Beijing 100049, China

---

[*] Corresponding author.

E-mail addresses: shankun@cigit.ac.cn (K. Shan); lrsong@ihb.ac.cn (L. Song)

ABSTRACT

Microcystis spp., which occur as colonies of different sizes under natural conditions, have expanded in temperate and tropical freshwater ecosystems and caused seriously environmental and ecological problems. In the current study, a Bayesian network (BN) framework was developed to access the probability of microcystins (MCs) risk in large shallow eutrophic lakes in China, namely, Taihu Lake, Chaohu Lake, and Dianchi Lake. By means of a knowledge-supported way, physicochemical factors, *Microcystis* morphospecies, and MCs were integrated into different network structures. The sensitive analysis illustrated that *Microcystis aeruginosa* biomass was overall the best predictor of MCs risk, and its high biomass relied on the combined condition that water temperature exceeded 24 °C and total phosphorus was above 0.2 mg/L. Simulated scenarios suggested that the probability of hazardous MCs (≥1.0 μg/L) was higher under interactive effect of temperature increase and nutrients (nitrogen and phosphorus) imbalance than that of warming alone. Likewise, data-driven model development using a naïve Bayes classifier and equal frequency discretization resulted in a substantial technical performance (CCI = 0.83, K = 0.60), but the performance significantly decreased when model excluded species-specific biomasses from input variables (CCI = 0.76, K = 0.40). The BN framework provided a useful screening tool to evaluate cyanotoxin in three studied lakes in China, and it can also be used in other lakes suffering from cyanobacterial blooms dominated by *Microcystis*.

Keywords: Bayesian networks; cyanobacterial blooms; *Microcystis* morphospecies; extracellular microcystin; temperature warming; nutrients reduction

1. Introduction

Microcystis is one of the most ubiquitous and notorious bloom-forming freshwater cyanobacterium and exhibits high phenotypic plasticity. Formation of colonies under natural conditions provides *Microcystis* a range of unique and highly adaptable eco-physiological traits, which enables them to dominate in eutrophic ecosystems associated with global warming (Carey et al., 2012; Harke et al., 2016; Zhang et al., 2006). In the current cyanobacteria taxonomic system, about 30 morphospecies of the genus of *Microcystis* have been recognised according to a variety of colonial morphologies, including irregular, sponge-like, spherical and elongated, and some show a visible margin (Komárek and Anagnostidis, 1999; Xiao et al., 2017). Six observed morphospecies, *M. aeruginosa* (Kützing) Kützing, *M. viridis* (A. Brown) Lemmermann, *M. flos-aquae* (Wittrock) Kirchner, *M. ichthyoblabe* Kützing, *M. novacekii* (Komárek) Compère, and *M. wesenbergii* (Komárek) Komárek, are regarded as the main species and dominate successively in water blooms (Xu et al., 2008). The triggering factors and mechanisms of colony-forming process in *Microcystis* populations have been the subject of past studies; however, the knowledge of succession pattern among different morphospecies in natural lakes remains unclear (Xiao et al., 2018; Zhu et al., 2016).

What is more, some species of *Microcystis* are potentially toxic and can produce highly stable and hepatotoxic polypeptides known as microcystins (MCs). The prevalence of toxic *Microcystis* blooms is emerging as one of the most important water quality and health issues in the world (Taranu et al., 2017). MCs can cause public health implications via a number of different exposure routes. Exposure to MC-LR (the most toxic microcystin variant) at sublethal doses have been reported to cause continual apoptotic cell death in the liver (Dong et al., 2016). MCs can remain in water environments due to their relative stability and half-life of days to weeks and enter food web by biotransformation and bioaccumulation, resulting in economic, social and ecological costs widely (Jones et al., 1995; Peng et al., 2010). The World Health Organization (WHO) proposed a widely used set of recommended action levels for risks associated with MCs exposure, which includes a safety limit of 1 μg/L MC-LR in drinking water and a chronic tolerable daily intake (TDI) of 0.04 $\mu g\ kg^{-1}$ body mass per day for human consumption (WHO, 1998). One approach for managing and reducing the risk of toxic *Microcystis* blooms is to identify environmental conditions under which high MCs concentrations are likely to occur (Yuan et al., 2014; Yuan and Pollard, 2017), but appropriate data to estimate widely applicable relationships are limited (Lürling et al., 2017).

The morphotypes and genotypes of *Microcystis* have been proven to be well-suited in indicating the dynamics of MC production (Le Ai Nguyen et al., 2012; Sabart et al., 2010; Srivastava et al., 2012; Wang et al., 2013). Considering the difference in environmental conditions (Joung et al., 2011; Rinta-Kanto et al., 2009), predictions by empirical models are complicated and likely site-specific. It is difficult to be generalised across relative large areas (Francy et al., 2016; Recknagel et al., 2017). Therefore, there is still a need of methodological approaches that can efficiently handle large and heterogeneous data, infer cause-and-effect relationships, and capture linear, non-linear, combinatorial, stochastic relationships among variables (Feki-Sahnoun et al., 2017).

Bayesian network (BN) represents a useful framework, because it can integrate multiple sources of information to estimate model parameter values and account for result uncertainty. As a result, BN can avoid reliance on a single deterministic outcome that does not reflect the inherent ecosystem variability (Arhonditsis et al., 2007). Other major benefits of the BN model include the capability to incorporate expert knowledge, automated learning of relationship structures, and the use of conditional probabilities (CPs) from databases with missing values (Landuyt et al., 2013).

These advantages have resulted in many applications of the BN model in managing eutrophic ecosystem. For instance, Borsuk et al. (2004) described BN-integrating models of the various processes involved in eutrophication in the Neuse River Estuary, North Carolina. The results of BN models gave decision makers a realistic appraisal of the chances of achieving desired outcomes under alternative nutrient reduction strategies. Hamilton et al. (2007) built a BN to determine the cyanobacterial bloom development in Deception Bay, Queensland, Australia. The most influential factors for *Lyngbya majuscula* bloom occurrence were identified to be water temperature, nutrient, and light availability. Rigosi et al. (2015) developed BN to determine the probability of cyanobacterial blooms with a broad range of temperature and nutrient inputs. The results suggested that the interactions between nutrients and temperature determined the high hazardous blooms. To the best of our knowledge, the BN model has not been applied in the domain of cyanotoxin prediction.

This paper aims to demonstrate the feasibility of the BN in predicting the risk of toxic *Microcystis* and MCs. Firstly, we compared succession pattern among *Microcystis*populations in different studied lakes, and identified the best variables for predicting total *Microcystis* biomass, toxic *Microcystis* biomass and MCs concentrations. Secondly, three knowledge-supported BN models were built to assess MCs risk under different scenarios. Finally, we compared the model performance of a purely data-driven BN including species-specific biomass with other common classification algorithms.

2. Materials and methods

2.1. Study site and data

In this study, we collected monthly data from three large shallow lakes between October 2008 and October 2010. All studied lakes have experienced severe eutrophication and increased occurrences of cyanobacterial blooms in the past few decades. Although it is a little out of date, cyanobacterial blooms in these lakes still reoccur and persist in recent years. Therefore, the data between 2008 and 2010 are valid for the model development. The number of sampling transects were 23 in Taihu Lake, 11 in Chaohu Lake, and 15 in Dianchi Lake (Fig. S1). Table 1 summarized the limnological characteristics of each lake.

Non-diazotrophic *Microcystis* was the most common bloom-forming genus and showed seasonal variability. In Taihu Lake and Chaohu Lake, the biomass of *Microcystis* grew rapidly and dominated cyanobacterial communities from late April, and then its biomass decreased gradually from November onwards. By contrast, *Microcystis* blooms in Dianchi Lake can persist for up to 10 months and cover the majority of the lake’s surface (Wu et al., 2016). *Microcystis* biomass was not uniformly distributed across the sampling sites. For example, the distribution of *Microcystis* biomass showed a decreasing trend from the western to the eastern region of Chaohu Lake, and declining from the northern to the southern region of Dianchi Lake. The northwestern sampling sites in Taihu Lake contributed a higher proportion of the total *Microcystis* biomass relative to other sampling sites.

During each transect, water samples were collected monthly at 0.5 m depth of surface layer. A litre of water was collected using a polymethyl methacrylate sampler and preserved with an acid Lugol's iodine solution (1% final concentration) to identify the phytoplanktonassemblages. Fresh samples were collected with a phytoplankton net (64 µm mesh size) from surface water (0 m to 0.5 m) and placed in 50 mL plastic bottles for identification of *Microcystis* morphospecies. A litre of water sample was brought back to the laboratory and filtered through GF/C glass microfiber filters (1.2 µm, Whatman) for nutrients and Chl-*α*analysis. Compared with relatively large data volume (n = 1090) of physical-chemical parameters and *Microcystis* morphospecies, MCs in water column were sampled across a total of 17 sampling transects. An additional litre of water sample was stored in glass bottles and frozen at −20 °C until MCs analysis.

Water quality parameters of each site, namely, water temperature (WT), electrical conductivity (EC), dissolved oxygen (DO), and pH were measured at 0.5 m depth of surface water with Multisensor Sonde (YSI 556MPS, USA). Transparency was measured with a 10 cm-diameter black-and-white Secchi disk (SD). Shade index (SI) was calculated as the ratio of lake depth to transparency. Nutrients were analysed for total nitrogen (TN), total phosphorus (TP), dissolved inorganic nitrogen (DIN; ammonium (N H4+) + nitrate (N O3-) + nitrite (N O2-)), and dissolved inorganic phosphorus (DIP). TP and TN concentrations were measured by colorimetry after digestion with $K_2S_2O_8$ + NaOH (Ebina et al., 1983). DIP was determined using the molybdenum blue method. N H4+ was determined using the indophenol blue method. N O3- and N O2- were measured with the cadmium reduction method (APHA, 1995).

The *Microcystis* morphospecies were classified according to the morphological descriptions given by Komárek and Komárková (2002). Cell density was estimated by picking each *Microcystis* morphotype colonies from fresh samples, fixing each with 0.1 mL of the Lugol's iodine solution and shaking it at 120 rpm until the colonies became unicellular to estimate cell density. Cells counting was repeated three times in a haemocytometer at 400 × power. The average number of cells in each type of colony was used to calculate the relative abundance of each type of *Microcystis* colony, which was further multiplied to the number of colonies present to estimate the abundance of each morphospecies. *Microcystis* biomass (B_M) was calculated on the basis of abundance (Biomass = algal density (1 g/cm^3) × size × abundance). More details are available in Hu et al. (2016) and Zhu et al. (2016). Species-specific biomasses regarding six commonly morphological subspecies included *Microcystis aeruginosa* (B_{MA}), *Microcystis wesenbergii* (B_{MW}), *Microcystis novacekii* (B_{MN}), *Microcystis ichthyoblabe* (B_{MI}), *Microcystis flos-aquae* (B_{MF}), and *Microcystis viridis* (B_{MV}).

Water samples (100 mL) were filtered through GF/C filters to remove plankton cells. The filtrates were measured by 96 wells filled with MCs for enzyme-linked immunosorbent assays (ELISA) (Wu et al., 2014). The samples were analysed in triplicate and compared with a 0.1 µg/L to 2.0 µg/L calibration curve of MC-LR standard (provided by the

Institute of Hydrobiology, Chinese Academy of Sciences) performed on each plate. Enzyme reactions were initiated by adding a substrate solution (0.1 M sodium acetate buffer with pH of 5.0) that contained 100 μg/mL of TMBZ and 0.005% (v/v) H_2O_2 and stopped with 1 M H_2SO_4. Absorbance at 450 nm was measured with a microtiter plate reader.

2.2. Bayesian network

The stepwise of the development of Bayesian network in this study was summarized in Fig. 1. The BN is probabilistic graphical model for a set of variables, which is defined in terms of two structural components, namely, directed acyclic graph (DAG) and conditional probabilitytables (CPTs) (Carvajal et al., 2015). The probability distribution of a node X is determined by the realised states of its preceding or parent nodes by using CPs $P\{X|\text{parents}[X]\}$. The joint probability distribution of all variables can be calculated to account for the independencies encoded by the network structure (Eq. [1]).

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|parents(X_i)) \quad (1)$$

where $P(X_1, X_2, \ldots, X_n)$ is the joint probability distribution of the variables $(X_1, X_2, \ldots, X_n)$. Variable X_i corresponds to a random variable represented by the node i in (1,…, n), and $parents(X_i)$ indicates a set of random variables associated with the parents of nodes i.

2.2.1. Knowledge-supported and data-driven BN development

Three network structures of knowledge-supported BN, which contained a two-layer structure and different variables, were constructed to integrate field data with empirical knowledge for managing hazardous *Microcystis* bloom. Variable discretisation and DAG construction were conducted following two stepwise approaches. Firstly, linear regression models with forward selection were used to explore the relationships amongst MCs, species-specific *Microcystis*biomass and environmental factors. Akaike information criterion (AIC) was selected as the coefficient of determination to identify the best model (Akaike, 1974). The discretization of continuous variables is a key step in the implementation of the BN model. Secondly, variable discretisation was performed on the basis of literatures and data distribution. MCs concentrations and the biomass of species-specific *Microcystis* can be discretised into three classes (low, moderate, and high) according to an alert level framework (ALF) for cyanobacteria in drinking water (Izydorczyk et al., 2009). WT was separated into three parts in accordance with distinct thresholds by Rigosi et al. (2015). By contrast, nutrients discretisation largely relied on their lower and upper quartile values (Table S1).

Alternatively, to develop a purely data-driven BN model, an exhaustive series of development settings were tested, including three type of network structure (na ïve Bayes (NB), tree-augmented na ïve (TAN), and augmented Bayesian na ïve with hill-climbing algorithms (BAN)), two type of discretization (equal width and equal frequency), and four discrete classes per variable (2–5 intervals). The principle of selecting variables depended on both the results of linear regressions and the low-cost availability in online monitoring. Models were developed based on each unique combination of settings. To select an optimal setting, the developed models were compared on the basis of predictive performance (Correctly Classified Instances, CCI). One hundred iterations were run, as exceeding iterations did not affect average model performance (Fig. 2). This model development process led to 2400 iterations, which were performed with the integration of three existing R packages ('bnlearn', 'infotheo', and 'caret') (Boets et al., 2015).

2.2.2. Model evaluation and scenario analysis

A sensitivity analysis was performed on the endpoint node to determine which *Microcystis*morphospecies most affected the assigned MCs concentrations (Marcot et al., 2006). The result of the sensitivity analysis generated percentages that were compared amongst nodes of the same network. This study assessed the effect of assigning a probability of 100% to a particular state of an input node by using the ‘enter finding’ Netica™ function while the distributions of the other input variables remained constant. The probabilities of *Microcystis*and toxic *M. aeruginosa* were evaluated under different combinations of WT, TP, and DIN concentrations. Scenarios, which included a trend of higher WT combined with TN decrease, were simulated to evaluate the risk of MCs concentrations. Changes in the probability of MCs hazard (low, moderate, and high) were calculated for all scenarios.

On the basis of a discretized dataset, our research randomised and partitioned the dataset into 10 equally sized sets. Twenty repetitions were performed by using a training set (9/10 of the data) and a testing set (1/10 of the data). The training set was used to learn the BN model. Two performance metrics, namely, Cohen’s Kappa (K) and CCI were adopted. The mean and standard deviations of CCI and K were calculated to evaluate the behavior of different BN models. In general, a model with a CCI of at least 0.70 and K higher than 0.4 is considered as a good model (Gabriels et al., 2007). In addition, seven common classification algorithms, namely, KNN (K-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), LR (logistic regression), C4.5, SVM (Support Vector Machine), RF (random forest), and MLP (Multi-layer Perceptron), were applied and compared for the same discretized data. Models were designed and evaluated using the Waikato Environment for Knowledge Analysis (WEKA) data mining software v.3.6.13 (Hall et al., 2009). Duncan’s test was performed in R package (agricolae) to test significant differences among predictive models.

3. Results

3.1. Relationship among environmental factors, species-specific biomasses and MCs

Linear regression models were used to determine the best variables for predicting species-specific biomasses and MCs. MCs concentrations across the lakes were explained by a combination of abiotic factors including SI, TN, and WT (Table 2, models 1–3). Despite the differences in regression coefficients, regression models for total *Microcystis* biomass (B_M) were obviously similar with those for *M. aeruginosa* biomass (B_{MA}). We found water temperature was one of the most explanatory variables in predicting *Microcystis* biomass. TP and DIN were the next important explanatory variables after temperature. The light conditions, indicated by the shade index, also played a significant role (Table 2, models 4–7). In comparison, the biomass of *Microcystis aeruginosa* was associated with TP, WT, DIN, and TN (Table 2, models 8–11).

The succession pattern among *Microcystis* populations in Taihu Lake and Chaohu Lake were similar (Fig. 3). *M. ichthyoblabe* and *M. flos-aquae* dominated in the early stages of bloom; subsequently, *M. aeruginosa* and *M. wesenbergii* began to co-dominate in summer and autumn, while *M. viridis* was often found in October and November. In Dianchi Lake, *M. viridis* and *M. wesenbergii* dominated the total *Microcytis* biomass, followed by *M. aeruginosa*, in comparable proportions from February to July. Sensitivity analysis was conducted with a four-node network including one additional parent node. Different *Microcystis* morphospecies that potentially affect the risk of MCs were connected one-at-a-time, and sensitivity results defined as percentage hazard were shown in Table 3. The results showed that *M. aeruginosa* biomass was the single most explanatory variable for evaluating the risk of MCs. As regards other morphospecies, their correlations with MCs relied on environmental conditions. For instance, the BN model including *M. viridis* showed that WT became the most important variable (network E in Table 3). Sensitivities of MCs hazard to the biomass of different morphospecies in the network were listed in the order of their importance: B_{MA}, B_{MV}, B_{MW}, B_{MI}, B_{MF}, and B_{MN}.

3.2. Predicting MC risk via knowledge-supported and data-driven BN models

Three knowledge-based BN were adopted. Firstly, a network of six nodes was adopted to analyse the relationships between total *Microcystis* biomass (B_M) with WT, TP, DIN, DIP and SI (Fig. 4a). The arrows in the network did not show the cause-effect relationships but diagnostic relationships, which provided a substantial prediction of B_M (CCI = 0.81 ± 0.04, K = 0.29 ± 0.04) (Table 4). Secondly, a network of five nodes was used to explore the relationships between toxic *M. aeruginosa* biomass (B_{MA}) with TP, WT, DIN, and TN (Fig. 4b). The current BN provided a stable and credible prediction of B_{MA}, because relatively few input variables reduced the amount of CPs (170) that had to be learned by filed observations. The predictive performance was good with a mean CCI of 0.94 ± 0.02 and mean K of 0.55 ± 0.05. Thirdly, a simplified network including four nodes, namely, MC hazard, WT, TN and SI, was employed (Fig. 4c). Although the predictive performance of MCs was relatively low (CCI = 0.62 ± 0.16, K = 0.14 ± 0.05), this BN model was easy to interpret and yielded important ecological insight between environmental variables and MCs.

A data-driven BN with 12 variables was developed to provide accurate prediction for MCs risk (Fig. 5). The optimal model development settings, defined via model development simulations, were equal width discretization and na ïve Bayes classifiers (Fig. 2). To reduce model complexity, discretization in two states was chosen as one of the optimal settings. The arrows in the network represented diagnostic relations among different input variables instead of causal ones. This BN model reduced the amount of conditional probabilities (CPs = 27), thereby leading to a substantial prediction of MCs (CCI = 0.83 ± 0.08, K = 0.60 ± 0.19). With the same discretized data, some comparative models including SVM and RF could also achieve the similar prediction accuracies even without parameters optimization (Table 5). When the biomasses of *M. aeruginosa* and *M. viridis* were included into the BN model, the predictive performance was obviously improved with the increase of CCI from 0.76 to 0.83. The importance of all input indicators in different lakes were sorted by sensitive analysis (Fig. 6). The results indicated that sensitivities of MCs hazard to the different input variables were listed in the order of their importance: B_{MA}, WT, B_{MV}, SI, N H4+, pH, DO, TP, B_M, N O3-, and TN. The risk of MCs in response to input variables in Dianchi Lake was clearly different from the other two lakes.

3.3. The BN model as a decision support tool to manage toxic **Microcystis**

Knowledge-supported BN allowed quantification of the probability of low, moderate, and high *Microcystis* biomass in given particular conditions of WT, TP, and DIN (Table 6). The high probability of total *Microcystis* biomass and *M. aeruginosa* biomass increased with increasing WT and TP concentrations by scenarios analysis. When combining WT and TP, probabilities varied, demonstrating an interaction rather than an addictive effect of these two factors. At low WT, the probability of high hazardous B_M was low even at high TP concentrations. At intermediate and high WT, there was evidence of dependency on TP for high hazardous B_M and B_{MA}. Moreover, when WT exceeded 24 °C and TP was above 0.2 mg/L, DIN concentration was the possible important explanatory variable. The highest hazardous B_{MA} was more likely to occur at lower DIN (< 0.4 mg/L).

To evaluate the effects of climate warming and TN decline, knowledge-supported BNs with endpoint of *M. aeruginosa* and MCs were employed (Fig. 4b and c). The decrease in TN from initial state to a state of low concentration ($TN < 2.0$ mg/L) slightly increased the probability of high MCs (≥ 1.0 μg/L) by 3%, but this change attenuated the probability of high B_{MA} by 1.1% (Table 7). Increasing WT from initial state to a warming state ($WT \geq 24$ °C) largely increased the probability of high B_{MA} and MCs by 14.9% and 12.8%, respectively. By contrast, when combining WT increase and TN reduction, the probability of high hazardous B_{MA} and MCs dramatically increased by 19.8% and 23.9%, respectively.

4. Discussion

Non-diazotrophic *Microcystis* populations in bloom season were composed of toxic and non-toxic strains. It is well recognized that quantitative PCR (qPCR) for MCs synthesis genes (*mcyA*-*mcyJ*) can well reflect the amount of MC-producing populations (Kardinaal et al., 2007; Singh et al., 2015). A small number of *Microcystis* strains, however, had been shown to contain the biosynthesis genes but lack detectable MCs (Mikalsen et al., 2003; Tillett et al., 2001). In addition, the time-consuming and skill-intensive characteristics of qPCR method made it difficult to be applied in on-site monitoring for water quality. *Microcystis*morphospecies, by contrast, were relatively easy to be identified and counted by microscopewith ImageTool software (i.e., Olympus DP-Soft, https://www.olympus-ims.com/), because of distinct phenotypes features for different *Microcystis* morphospecies. For instance, toxic *M. aeruginosa* generally grows irregularly in shape, elongated or lobed containing distinct holes, relatively firm. Non-toxic *M. wesenbergii* is irregular, spheroidal to lobate or elongate with a visible margin that is filled with mucilage (Komárek and Komárková, 2002).

Owing to the diversity of MC genes within toxic *Microcystis* populations, it is still unclear to which extend morphological characteristics are linked to the ability to produce MCs (Kurmayer et al., 2002; Via-Ordorika et al., 2004). In this study, although succession patterns of *Microcystis* populations in three studied lakes were different, the microscopical analysis of different morphospecies was considered indicative for MCs production. Sensitivity analysis with a four-node BN showed that *M. aeruginosa* was the most important morphotypes in determining the high MCs. The results echoed the previous experiments in Taihu Lake by Otten and Paerl (2011), which suggested that caution should be exercised when the *M. aeruginosa* morphotype is present. *M.*

viridis was the next important biotic indicators, but MCs production of this genus was dependent on environmental conditions (Ozawa et al., 2005; Song et al., 1998). *M. wesenbergii* so far has never been found to produce microcystin, while its correlation with MCs might be attributed to the coexistence with *M. aeruginosa*. The morphospecies *M. flos-aquae* and *M. ichthyoblabe* were generally considered capable of MCs production, but both appeared to be responsible for a small portion of the toxic potential. In agreement with results from laboratory cultures (Hu et al., 2016), strains of *M. viridis* from Dianchi Lake produced large amounts of MCs, whereas strains of *M. flos-aquae* from Taihu Lake produced trace amounts of MCs or no toxin.

Climate warming has been considered the key factor responsible for promoting cyanobacterial bloom (Kosten et al., 2012; O'neil et al., 2012; Paerl and Paul, 2012). More recent studies also suggested that warming temperature could prompt the growth of some *Microcystis* strains, potentially changing the distribution of cyanotoxin (Bui et al., 2018; Mantzouki et al., 2018). In shallow well-mixed lake, climate warming increases nutrients concentrations by enhancing mineralization and anoxia-mediated sediment phosphorus release to sustain the nutrients requirement of cyanobacteria (Jensen and Andersen, 1992). Long-term trends of nutrients provided evidence that the gradually decline TN:TP over the last decade in some eutrophic lakes, i.e., Taihu Lake and Dianchi Lake (Xu et al., 2017; Yan et al., 2016). Therefore, the level of nutrient interaction necessary to counteract the effects of warming and how nutrients interact with temperature should be thoroughly understood to better manage the risk of cyanotoxins. The results of BNs corroborated previous finding in experimental manipulations that higher temperature coupled with elevated phosphorus concentrations promote growth rates of toxic *Microcystis* cell (Davis et al., 2009). Interestingly, when water temperature exceeded 24 °C and total phosphorus was above 0.2 mg/L, evidence from BNs showed a coincidental correspondence between low DIN concentrations (< 0.4 mg/L) and high biomass of toxic *M. aeruginosa.* High DIN concentrations (> 1.5 mg/L), by contrast, corresponded to high biomass of total *Microcystis*and considerable biomass of *M. aeruginosa.* Even though it was difficult to distinguish between cause and effect, prior work has shown a remarkable agreement that nitrogen concentration and more specifically the availability of different nitrogen forms may influence the overall toxicity of blooms (Donald et al., 2011).

Nutrient addition bioassays in Taihu Lake indicated that the lake shifts from phosphorus-limitation in winter-spring to nitrogen-limitation in cyanobacteria-dominated summer and fall months (Paerl et al., 2011). During summer months, regenerated nitrogen (N) was the dominant source of combined N available to phytoplankton (Paerl et al., 2015). Meanwhile, *Microcystis* appear to out-compete diazotrophic competitors, *Dolichospermum* and *Aphanizomenon*, for dwindling N sources (Blomqvist et al., 1994), including ammonium generated from decomposing organic matter (N regeneration). Our results indicated that ammonium, rather than nitrate, was the most important N forms in predicting MCs, buttressing previous finding by Monchamp et al. (2014). Beversdorf et al. (2013) proposed that a new N production was a limiting factor to support the growth of the potentially toxic *Microcystis*. Thus, in the summer, the low concentrations of DIN may be associated with high biomass of *M. aeruginosa.* Furthermore, simulated scenarios of TN decrease also predicted a lot increase in MC concentrations, but toxic *M. aeruginosa* biomass tended to be constant.

This result lent support to field observations in Taihu that decreasing trend of TN does not coincide with distinctive changes in cyanobacterial toxicity (Hu et al., 2016). More seriously, we predicted that MC risk was much higher under the synergistic effect between higher temperature and nutrients (TN and TP) imbalance, as the predicted further warming of at least 1.8 °C until the end of this century (Wagner and Adrian, 2009).

The concentrations of MCs in water column are a function of the biomass of toxigenic species and the amount of toxin per unit cell. While MC cell quotas are not only determined by the presence of the synthesizing genes, but also by environmental factors that influence MC synthesis or fate (Pacheco et al., 2016). Therefore, prediction of the variations in MCs should require a knowledge of environmental conditions under which certain *Microcystis*species are more likely to produce toxins. Here, we explored a framework of BN for predicting the likelihood of high MCs concentrations. The predictive performance was substantial based on twelve biotic and abiotic variables. However, if the proposed model will be adopted in the real-time warning system, a temporal discrepancy from strain growth to toxin release should be deliberated. The inputs and outputs in predictive model should not be synchronized variables, because most MCs are usually released into water after the lysisof toxic *Microcystis* cell (Dong et al., 2016). This problem, however, is unlikely to be possible solved under the current framework of BN. Other technology such as dynamic Bayesian network would be more suitable and worth further study.

Real-world ecosystems will differ in many aspects, including temperature range, nutrients concentrations, cyanobacterial communities, and succession pattern. As expected, management and prediction strategies for cyanotoxins were largely dependent on lake-specific environmental gradients (Taranu et al., 2012). Hence, the repetition of this model development process in other lakes suffering from *Microcystis* blooms may be preferable to the indiscriminate use of models. The efficacy of different management options can be tested via the built BN. Same as many studies, this study has limitations and several open issues should be considered. Firstly, there was no strictly independence assumption on input variables of naive Bayes classifier, because this methods was prepared to investigate the possible of applying date-driven BN. Secondly, are there some other low-cost factors can be used to achieve MCs prediction? In future, continuously meteorological and hydrologic factors will be added to improve the predictive performance of BN. Lastly, although data in this study were collected during the period of 2008–2010, they are still valid to construct and test BNs models in these three lakes where no marked change of phytoplankton and nutrients happened in the last decade. The future studies including more recent data will certainly further improve our understanding and modelling of MCs risk in China and other countries.

5. Conclusions

This study highlighted the importance of succession pattern of *Microcystis* morphospecies in MCs risk assessment, and provided a reference for the prediction of cyanotoxin. The proposed BNs provided an objective, systematic and applicable way of estimating potentially toxic ‘hot spots’ of unacceptable MCs levels in three studied lakes. The framework of BN herein allowed us to identify the importance of explanatory variables, and to test the efficacy of different management options via scenario analysis. We suggested that future nutrients reduction should consider toxic versus non-toxic cyanobacterial populations dynamics and the potential impacts from climate warming.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 51609229; No. 41701247), Chongqing Science and Technology Commission (No. cstc2017jcyjAX0241) and National Key Scientific and Technological Project of China (2014ZX07104-006). The field study in three studied lakes of China were financed by National Basic Research Program of China (2008CB418006). Special thanks also to the help from Dr. Wei Chen, Dr. Yanlong Wu, Dr. Liang Peng, Dr. Liming Liu, Dr. Qichao Zhou, Dr. Yunlu Jia and Dr. Lili Hu.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control 19, 716-723.

Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., Reckhow, K.H., 2007. Eutrophication risk assessment using Bayesian calibration of process-based models: application to a mesotrophic lake. Ecological modelling 208, 215-229.

Beversdorf, L.J., Miller, T.R., McMahon, K.D., 2013. The role of nitrogen fixation in cyanobacterial bloom toxicity in a temperate, eutrophic lake. PLoS One 8, e56103.

Boets, P., Landuyt, D., Everaert, G., Broekx, S., Goethals, P.L., 2015. Evaluation and comparison of data-driven and knowledge-supported Bayesian Belief Networks to assess the habitat suitability for alien macroinvertebrates. Environmental Modelling & Software 74, 92-103.

Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173, 219-239.

Burson, A., Stomp, M., Akil, L., Brussaard, C.P., Huisman, J., 2016. Unbalanced reduction of nutrient loads has created an offshore gradient from phosphorus to nitrogen limitation in the North Sea. Limnology and Oceanography 61, 869-888.

Carey, C.C., Ibelings, B.W., Hoffmann, E.P., Hamilton, D.P., Brookes, J.D., 2012. Eco-physiological adaptations that favor freshwater cyanobacteria in a changing climate. Water research 46, 1394-1407.

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

Dong, X., Zeng, S., Bai, F., Li, D., He, M., 2016. Extracellular microcystin prediction based on toxigenic *Microcystis* detection in a eutrophic lake. Scientific reports 6, 1-8.

Gabriels, W., Goethals, P.L., Dedecker, A.P., Lek, S., De Pauw, N., 2007. Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. Aquatic Ecology 41, 427-441.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter 11, 10-18.

Hamilton, G.S., Fielding, F., Chiffings, A.W., Hart, B.T., Johnstone, R.W., Mengersen, K., 2007. Investigating the use of a Bayesian network to model the risk of *Lyngbya majuscula* bloom initiation in Deception Bay, Queensland, Australia. Human and Ecological Risk Assessment 13, 1271-1287.

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.

Hu, L., Shan, K., Lin, L., Shen, W., Huang, L., Gan, N., Song, L., 2016. Multi-year assessment of toxic genotypes and microcystin concentration in northern lake Taihu, China. Toxins 8, 23.

Izydorczyk, K., Carpentier, C., Mrówczyński, J., Wagenvoort, A., Jurczak, T., Tarczyńska, M., 2009. Establishment of an Alert Level Framework for cyanobacteria in drinking water resources by using the Algae Online Analyser for monitoring cyanobacterial chlorophyll a. Water research 43, 989-996.

Jensen, H.S., Andersen, F.O., 1992. Importance of temperature, nitrate, and pH for phosphate release from aerobic sediments of four shallow, eutrophic lakes. Limnology and Oceanography 37, 577-589.

Jones, G.J., Falconer, I.R., Wilkins, R.M., 1995. Persistence of cyclic peptide toxins in dried *Microcystis aeruginosa* crusts from Lake Mokoan, Australia. Environmental toxicology and water quality 10, 19-24.

Joung, S.H., Oh, H.M., Ko, S.R., Ahn, C.Y., 2011. Correlations between environmental factors and toxic and non-toxic *Microcystis* dynamics during bloom in Daechung Reservoir, Korea. Harmful Algae 10, 188-193.

Kardinaal, W.E.A., Janse, I., Kamst-van Agterveld, M., Meima, M., Snoek, J., Mur, L. R., Huisman, J., Zwart, G., Visser, P.M., 2007. *Microcystis* genotype succession in relation to microcystin concentrations in freshwater lakes. Aquatic Microbial Ecology 48, 1-12.

Komárek, J., Komárková, J., 2002. Review of the European *Microcystis*-morphospecies (Cyanoprokaryotes) from nature. Czech Phycology, Olomouc 2, 1-24.

Komárek, J., Anagnostidis, K., 1999. Cyanoprokaryota. In: Ettl, H., Gardner, G., Heyning, H., Mollenheuer, D. (Eds.), 1: Chroococcales. Süsswasserflora von Mitteleurope. Gustav Fischer, pp. 225-236.

Kosten, S., Huszar, V.L., Bécares, E., Costa, L.S., Donk, E., Hansson, L.A., Jeppesen, E., Kruk, C., Lacerot, G., Mazzeo, N., De Meester, L., Moss, B., Lürling, M., Nõges, T., Romo, S., Scheffer, M., 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Global Change Biology 18, 118-126.

Landuyt, D., Broekx, S., D'hondt, R., Engelen, G., Aertsens, J., Goethals, P.L., 2013. A review of Bayesian belief networks in ecosystem service modelling. Environmental Modelling & Software 46, 1-11.

Le Ai Nguyen, V., Tanabe, Y., Matsuura, H., Kaya, K., Watanabe, M.M., 2012. Morphological, biochemical and phylogenetic assessments of water-bloom-forming tropical morphospecies of *Microcystis* (Chroococcales, Cyanobacteria). Phycological research 60, 208-222.

Lürling, M., Eshetu, F., Faassen, E.J., Kosten, S., Huszar, V. L., 2013. Comparison of cyanobacterial and green algal growth rates at different temperatures. Freshwater Biology 58, 552-559.

Lürling, M., Van Oosterhout, F., Faassen, E., 2017. Eutrophication and warming boost cyanobacterial biomass and microcystins. Toxins 9, 64.

Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Canadian Journal of Forest Research 36, 3063-3074.

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.

Otten, T.G., Xu, H., Qin, B., Zhu, G., Paerl, H.W., 2012. Spatiotemporal patterns and ecophysiology of toxigenic *Microcystis* blooms in Lake Taihu, China: implications for water quality management. Environmental science & technology 46, 3480-3488.

Ozawa, K., Fujioka, H., Muranaka, M., Yokoyama, A., Katagami, Y., Homma, T., et al., 2005. Spatial distribution and temporal variation of *Microcystis* species composition and microcystin concentration in Lake Biwa. Environmental toxicology 20, 270-276.

Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. Water research 46, 1349-1363.

Peng, L., Liu, Y., Chen, W., Liu, L., Kent, M., Song, L., 2010. Health risks associated with consumption of microcystin-contaminated fish and shellfish in three Chinese lakes: significance for freshwater aquacultures. Ecotoxicology and environmental safety 73, 1804-1811.

Recknagel, F., Orr, P.T., Bartkow, M., Swanepoel, A., Cao, H., 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. Harmful algae 69, 18-27.

Rigosi, A., Hanson, P., Hamilton, D.P., Hipsey, M., 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. Ecological Applications 25, 186-199.

Rinta-Kanto, J.M., Konopko, E.A., DeBruyn, J.M., Bourbonniere, R.A., Boyer, G.L., Wilhelm, S.W., 2009. Lake Erie *Microcystis*: relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. Harmful Algae 8, 665-673.

Sabart, M., Pobel, D., Briand, E., Combourieu, B., Salencon, M.J., Humbert, J.F., Latour, D., 2010. Spatiotemporal variations in microcystin concentrations and in the proportions of microcystin-producing cells in several *Microcystis aeruginosa* populations. Applied and environmental microbiology 76 (14), 4750-4759.

Singh, S., Rai, P.K., Chau, R., Ravi, A.K., Neilan, B.A., Asthana, R.K., 2015. Temporal variations in microcystin-producing cells and microcystin concentrations in two fresh water ponds. Water research 69, 131-142.

Srivastava, A., Choi, G.G., Ahn, C.Y., Oh, H.M., Ravi, A.K., Asthana, R.K., 2012. Dynamics of microcystin production and quantification of potentially toxigenic *Microcystis* sp. using real-time PCR. Water research 46, 817-827.

Song, L., Sano, T., Li, R., Watanabe, M.M., Liu, Y., Kaya, K., 1998. Microcystin production of *Microcystis viridis* (cyanobacteria) under different culture conditions. Phycological research 46, 19-23.

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 Biology 18, 3477-3490.

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 Ecology and Biogeography 26, 625-637.

Wang, X., Sun, M., Xie, M., Liu, M., Luo, L., Li, P., Kong, F., 2013. Differences in microcystin production and genotype composition among *Microcystis* colonies of different sizes in Lake Taihu. Water research 47, 5659-5669.

World Health Organization (WHO), 1998. World Health Organization (WHO) Guidelines for Drinking Water Quality, vol. 1, second ed., Recommendations /World Health Organization, World Health Organization, Geneva, Switzerland (1998) 36 pp.

Wu, Y., Li, L., Gan, N., Zheng, L., Ma, H., Shan, K., Liu, J., Xiao, B., Song, L., 2014. Seasonal dynamics of water bloom-forming *Microcystis* morphospecies and the associated extracellular microcystin concentrations in large, shallow, eutrophic Dianchi Lake. Journal of Environmental Sciences 26, 1921-1929.

Wu, Y., Li, L., Zheng, L., Dai, G., Ma, H., Shan, K., Wu, H., Zhou, Q., Song L., 2016. Patterns of succession between bloom-forming cyanobacteria *Aphanizomenon flos-aquae* and *Microcystis* and related environmental factors in large, shallow Dianchi Lake, China. Hydrobiologia 765, 1-13.

Xiao, M., Willis, A., Burford, M.A., Li, M., 2017. A meta-analysis comparing cell-division and cell-adhesion in Microcystis colony formation. Harmful algae 67, 85-91.

Xiao, M., Li, M., Reynolds, C.S., 2018. Colony formation in the cyanobacterium *Microcystis*. Biological Reviews. https://doi.org/10.1111/brv.12401.

Xu, H., Paerl, H.W., Zhu, G., Qin, B., Hall, N.S., Zhu, M., 2017. Long-term nutrient trends and harmful cyanobacterial bloom potential in hypertrophic Lake Taihu, China. Hydrobiologia 787, 229-242.

Xu, Y., Wu, Z., Yu, B., Peng, X., Yu, G., Wei, Z., Wang, G., Li, R., 2008. Non-microcystin producing *Microcystis wesenbergii* (Komárek) Komárek (Cyanobacteria) representing a main waterbloom-forming species in Chinese waters. Environmental Pollution 156, 162-167.

Yan, Z., Han, W., Peñuelas, J., Sardans, J., Elser, J.J., Du, E., Reich, P.B., Fang J., 2016. Phosphorus accumulates faster than nitrogen globally in freshwater ecosystems under anthropogenic impacts. Ecology Letters 19, 1237-1246.

Yang, H., Flower, R.J., Thompson, J.R., 2013. Sustaining China's Water Resources. Science 339, 141-141.

Yang, H., Huang, X., Thompson, J.R., Flower, R.J., 2015. Enforcement key to China's environment. Science 347, 834-835.

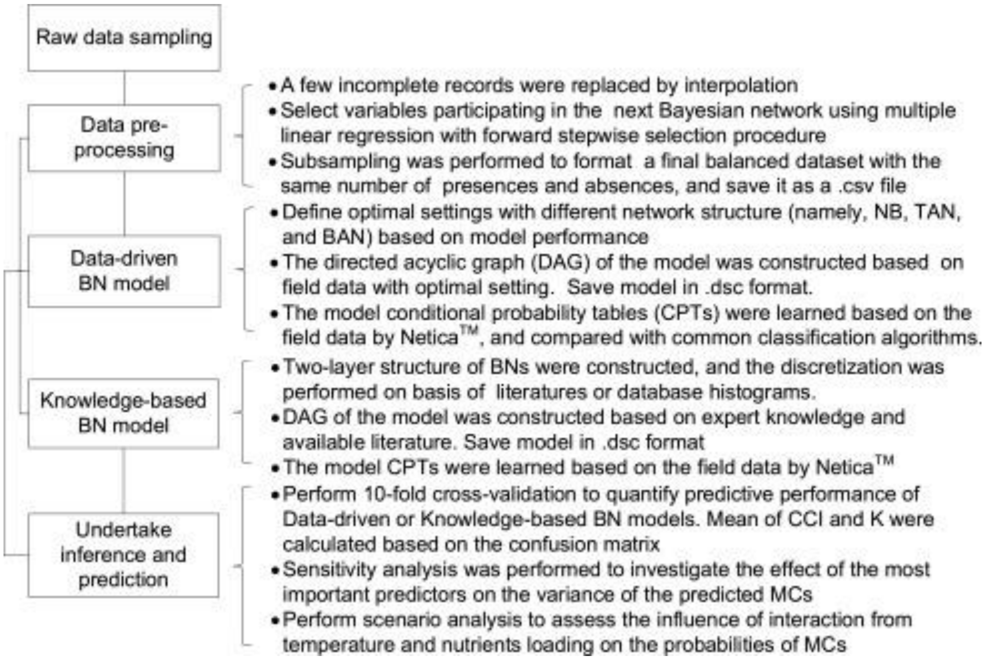
Yang, H., Xie, P., Ni, L., Flower, R.J., 2012. Pollution in the Yangtze. Science 337, 410-410.

Yuan, L.L., Pollard, A.I., Pather, S., Oliver, J.L., D'Anglada, L., 2014. Managing microcystin: identifying national-scale thresholds for total nitrogen and chlorophyll a. Freshwater biology 59, 1970-1981.

Yuan, L.L., Pollard, A.I., 2017. Using National-scale data to develop nutrient – microcystin relationships that guide management decisions. Environmental science & technology 51 (12), 6972-6980.

Zhang, M., Xie, P., Xu, J., Liu, B., Yang, H., 2006. Spatiotemporal variations of internal P-loading and the related mechanisms in the large shallow Lake Chaohu. Science in China Series D: Earth Sciences 49, 72-81.

Zhu, W., Zhou, X., Chen, H., Gao, L., Xiao, M., Li, M., 2016. High nutrient concentration and temperature alleviated formation of large colonies of *Microcystis*: evidence from field investigations and laboratory experiments. Water research 101, 167-175.



- A few incomplete records were replaced by interpolation
- Select variables participating in the next Bayesian network using multiple linear regression with forward stepwise selection procedure
- Subsampling was performed to format a final balanced dataset with the same number of presences and absences, and save it as a .csv file

- Define optimal settings with different network structure (namely, NB, TAN, and BAN) based on model performance
- The directed acyclic graph (DAG) of the model was constructed based on field data with optimal setting. Save model in .dsc format.
- The model conditional probability tables (CPTs) were learned based on the field data by Netica^{TM}, and compared with common classification algorithms.

- Two-layer structure of BNs were constructed, and the discretization was performed on basis of literatures or database histograms.
- DAG of the model was constructed based on expert knowledge and available literature. Save model in .dsc format
- The model CPTs were learned based on the field data by Netica^{TM}

- Perform 10-fold cross-validation to quantify predictive performance of Data-driven or Knowledge-based BN models. Mean of CCI and K were calculated based on the confusion matrix
- Sensitivity analysis was performed to investigate the effect of the most important predictors on the variance of the predicted MCs
- Perform scenario analysis to assess the influence of interaction from temperature and nutrients loading on the probabilities of MCs

Fig. 1. Flowchart of the BN development for assessing MCs risk by data-driven and knowledge-supported ways.

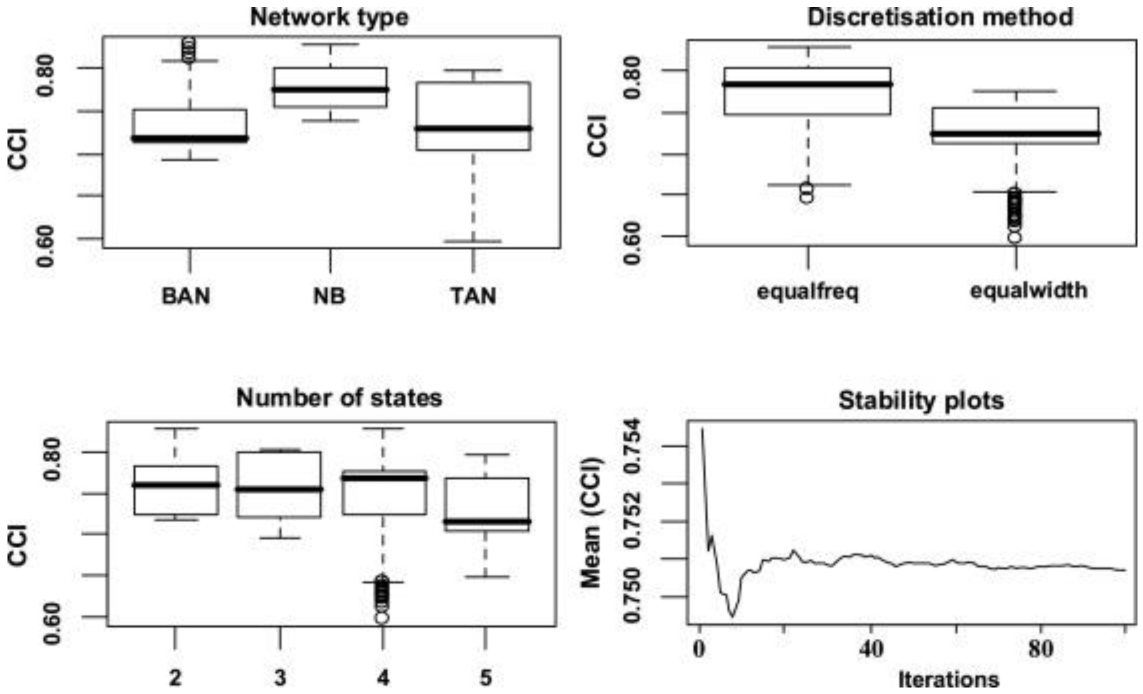


Fig. 2. The effect of parameter selection on predictive performance of date-driven BBN models. Three types of network structure were tested including naïve Bayes (NB), tree-augmented naïve (TAN), augmented Bayesian naïve with hill-climbing algorithms (BAN).

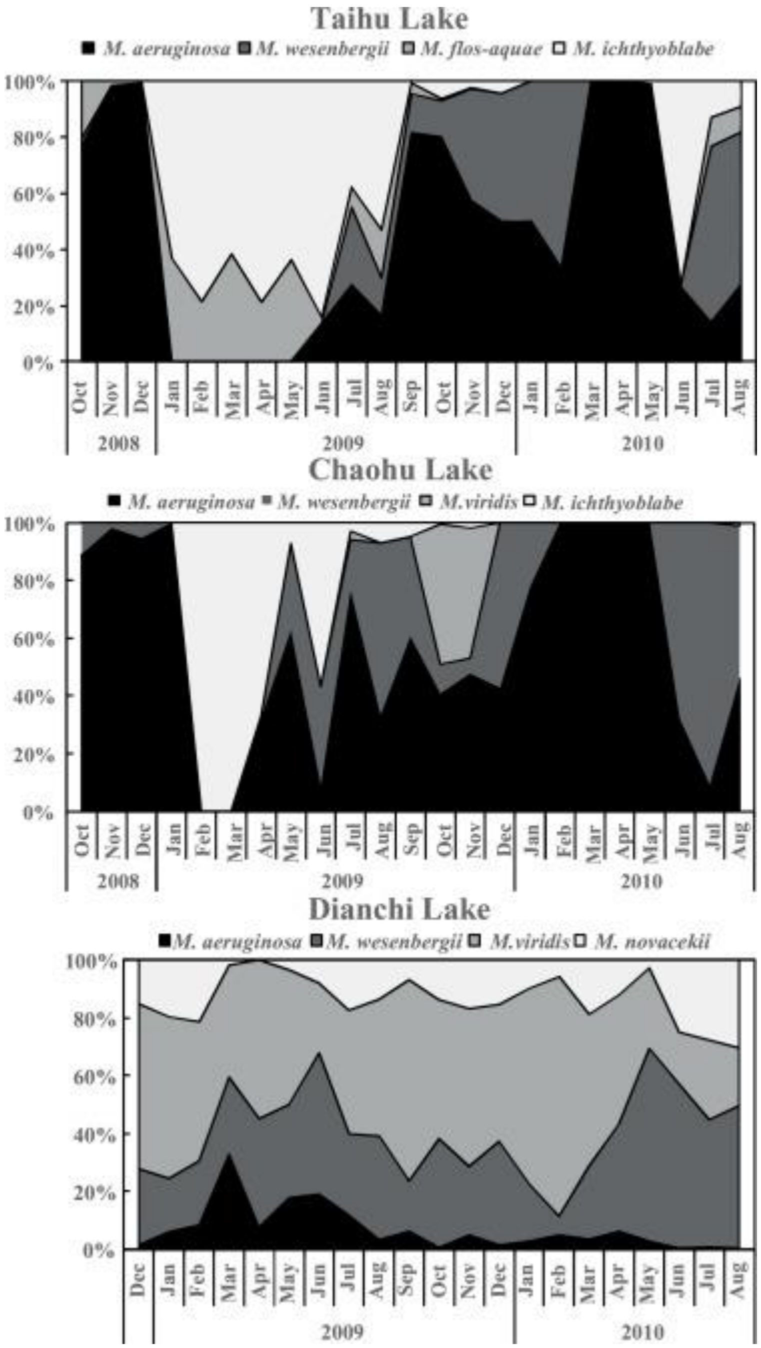


Fig. 3. Seasonal variations in the proportion of various *Microcystis* species based on colonial morphology. The data are average values at multiple sites in each lake.

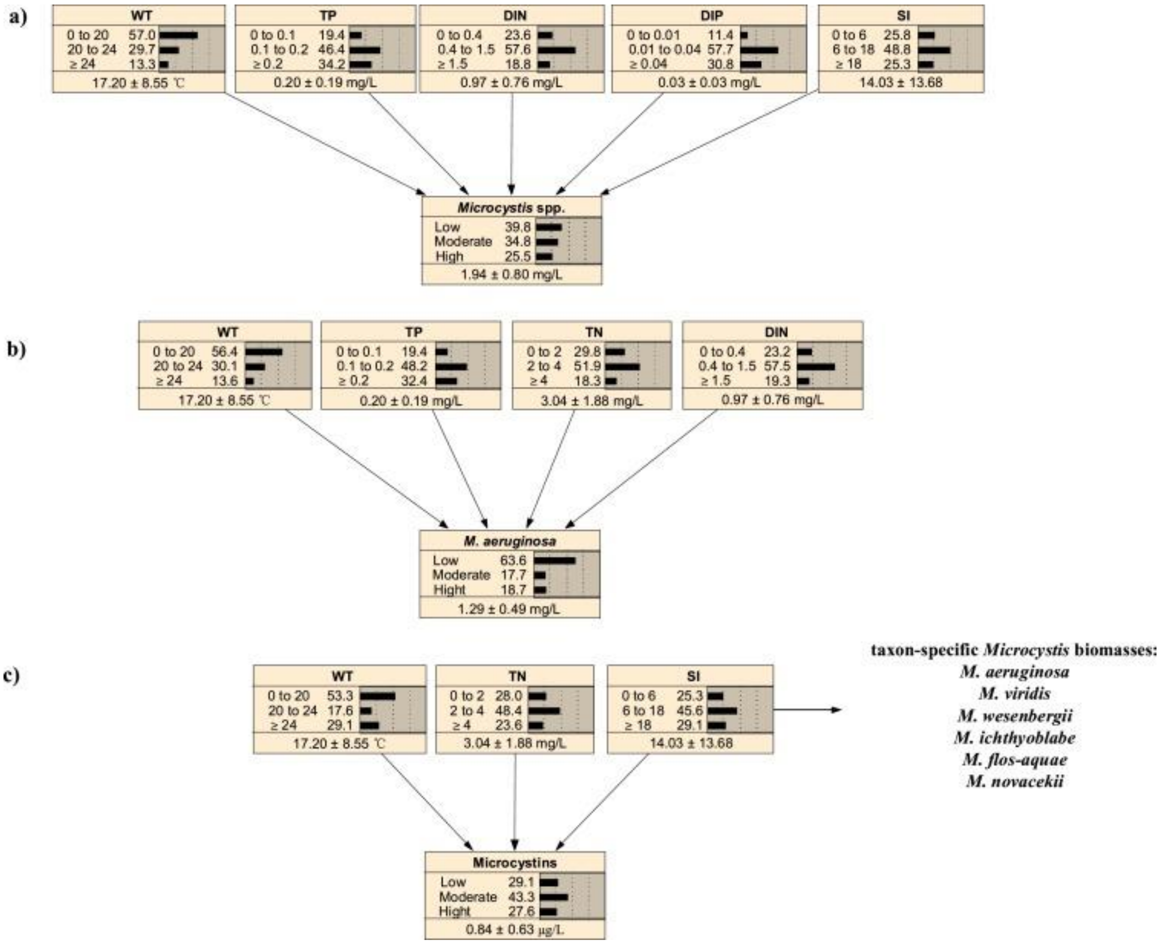


Fig. 4. Bayesian network structure for assessing toxic cyanobacteria hazard class represented as (a) total biomass of *Microcystis*, B_M, (b) *Microcystis aeruginosa* biomass, $B_{MA,}$ and (c) microcystins concentration, MCs. All continuous variables are discretized into three states, and the black horizontal bars along each node indicate the probability (%) of being in a particular state. In panel c, the arrow indicates that network structure added alternatively with different taxon-specific *Microcystis* biomasses (*M. aeruginosa*, *M. viridis*, *M. wesenbergii*, *M. ichthyoblabe*, *M. flos-aquae*, *M. novacekii*).

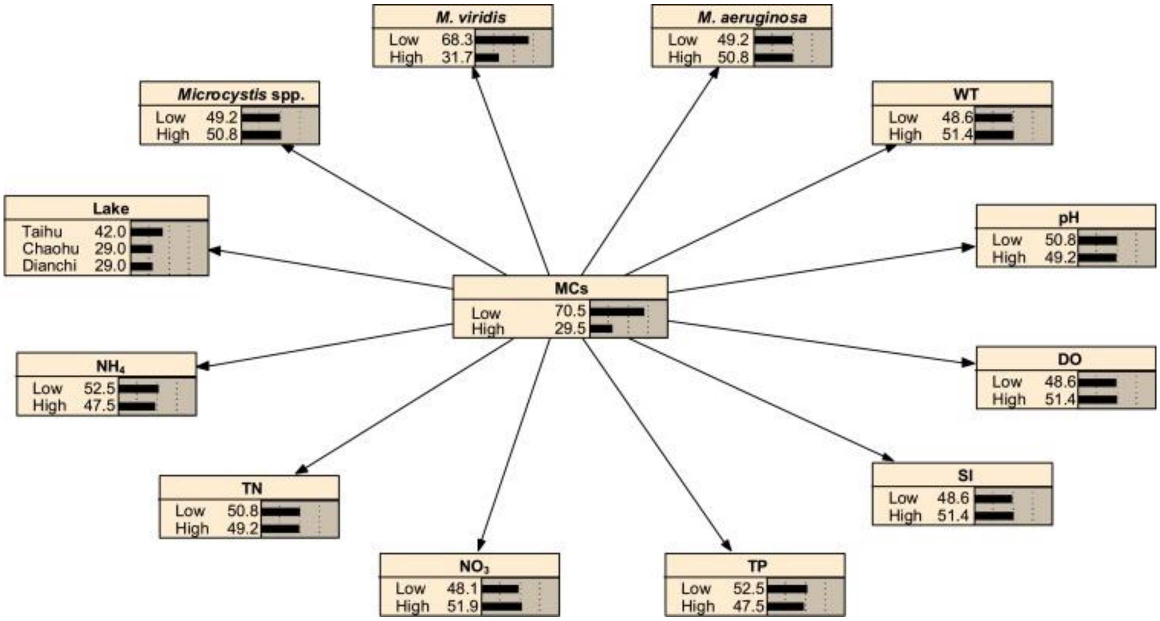


Fig. 5. BN created from data-driven model development. Apart from specific variable ‘Lake’, which refers to the three sampled lakes, all continuous variables are discretized into two states, and the network structure corresponds to that of a Naïve Bayes classifier. The black horizontal bars within each node indicate the probability (%) of being in a particular state.

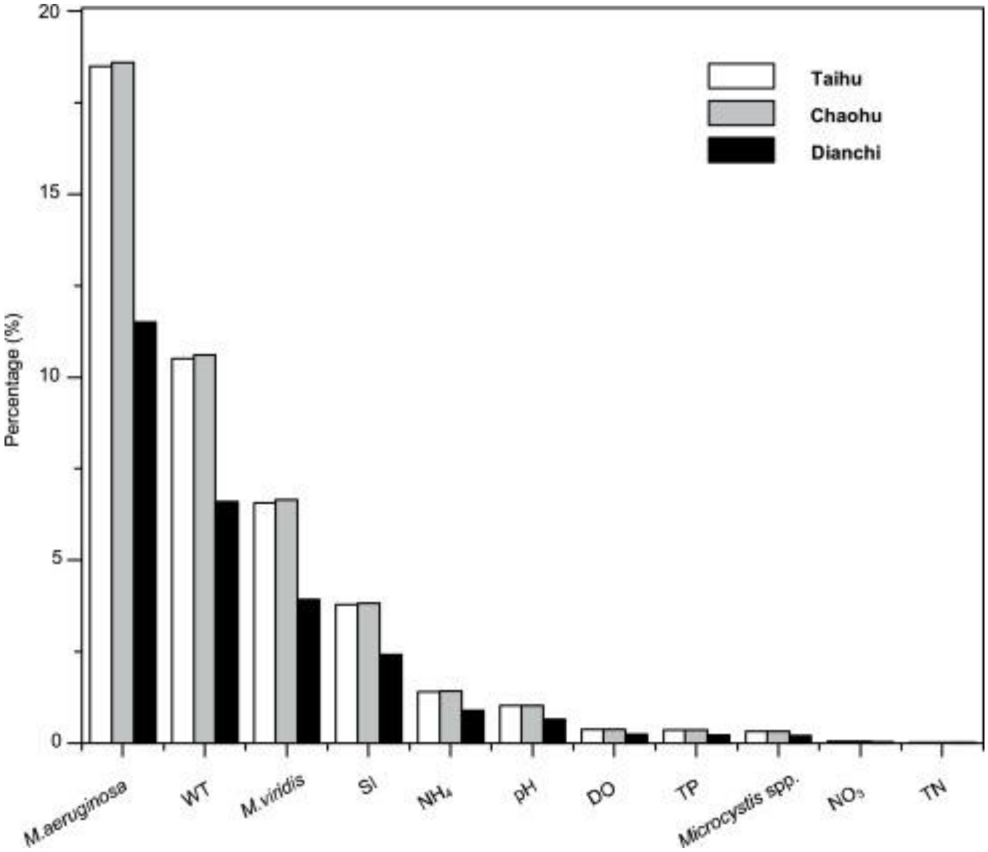


Fig. 6. Sensitivities of MCs hazard to different biotic and abiotic factors in three studied lakes, adopting the data-driven BN.

Table 1. Summary of the average of limnological variables and the results of their correlations with microcystin concentrations in three studied lakes during the period 2008–2010.

	Taihu Lake	**Chaohu Lake**	**Dianchi Lake**
Latitude	30°56′˜31°34′N	30°25′˜31°43′N	24°29′˜25°28′N
Longitude	119°53′˜120°34′E	117°17′˜117°52′E	102°29′˜103°01′E
Altitude (m)	3	170	1886
Mean depth (m)	1.9	3.1	4.7
Surface area (km^2)	2338	780	306
Water residence time (year)	1	0.5	3.5
pH	8.4 ± 0.6	8.4 ± 0.4	9.2 ± 0.7 (+**)
DO (mg/L)	8.9 ± 2.4 (–**)	8.7 ± 2.6	8.0 ± 3.5
SD (cm)	30 ± 16 (+*)	30 ± 13 (–**)	20 ± 10
Wind speed (m/s)	3.1 ± 1.8	2.9 ± 1.7 (–*)	3.4 ± 1.9 (–*)
Chl *α* (μg/L)	44.7 ± 73.2 (+**)	60.5 ± 117.8	157.2 ± 203.3 (+**)
TN (mg/L)	3.56 ± 2.68 (+**)	2.64 ± 1.51	3.51 ± 2.19
DIN (mg/L)	1.46 ± 0.69	1.04 ± 0.62 (–*)	0.68 ± 0.71 (–**)
TP (mg/L)	0.16 ± 0.09	0.18 ± 0.23	0.21 ± 0.20
DIP (mg/L)	0.02 ± 0.01	0.03 ± 0.02	0.04 ± 0.05
Cyanobacteria biomass (mg/L)	17.8 ± 34.4 (+**)	34.8 ± 83.7 (+**)	54.6 ± 52.8 (+*)
***Microcystis* biomass (mg/L)**	16.7 ± 33.7 (+**)	22.4 ± 80.3 (+**)	41.8 ± 36.8 (+**)
***Dolichospermum* biomass (mg/L)**	0.9 ± 2.4 (+*)	10.6 ± 18.3	3.7 ± 10.2
***Aphanizomenon* biomass (mg/L)**	0.1 ± 0.2	1.4 ± 3.3	8.6 ± 30.5
Dissolved microcystin (μg/L)	0.97 ± 0.64	1.02 ± 0.75	0.48 ± 0.24

Note: Signs within brackets behind each variable indicate this variable either promotes (+) or reduces (–) the MCs concentration in lake-specific dataset.* Correlation with MCs is significant at $P < 0.05$. ** Correlation with MCs is significant at $P < 0.01$.

Table 2. Linear regression models explaining the toxicity of cyanobacterial blooms. The model results come from a forward stepwise selection procedure, and explain microcystin concentrations (MCs), total *Microcystis* biomass (B_M) and *Microcystis aeruginosa* biomass (B_{MA}) by several selected variables, including water temperature (WT), total nitrogen (TN), total phosphorus (TP), dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), and shade index (SI). All regression models were significant ($P < 0.0001$).

No.	**Dependent variable**	**Linear model**	r^2_{adj}	**AIC**	***F***	**df**
1		$-0.29\mathrm{LogSI}^{***}$	0.08	126	19	201
2	Log MCs	$-0.31\mathrm{LogSI}^{***} + 0.25\ \mathrm{Log\ TN}^{**}$	0.12	119	14	200
3		$-0.31\mathrm{LogSI}^{***} + 0.23\ \mathrm{Log\ TN}^{**} + 0.19\ \mathrm{Log\ WT}^{*}$	0.14	115	12	199
4		$-0.47^{***} + 1.04\ \mathrm{LogWT}^{***}$	0.19	2098	249	1088
5		$0.38^{***} + 1.05\mathrm{LogWT}^{***} + 1.06\ \mathrm{LogTP}^{***}$	0.36	1835	308	1087
6	Log B_M	$0.36^{***} + 1.01\mathrm{LogWT}^{***} + 1.07\ \mathrm{LogTP}^{***} - 0.43\mathrm{Log\ DIN}^{***}$	0.42	1732	263	1086
7		$-0.53^{***} + 0.98\mathrm{LogWT}^{***} + 0.82\mathrm{LogTP}^{***} - 0.33\mathrm{LogDIN}^{***} + 0.45\mathrm{LogSI}^{***} - 0.11\mathrm{LogDIP}^{**}$	0.46	1654	156	1083
8	Log B_{MA}	$0.46^{***} + 0.35\mathrm{LogTP}^{***}$	0.09	508	108	1088
9		$0.18^{***} + 0.35\mathrm{LogTP}^{***} + 0.24\mathrm{LogWT}^{***}$	0.14	450	88	1087
10		$0.18^{***} + 0.35\mathrm{LogTP}^{***} + 0.24\mathrm{LogWT}^{***} - 0.06\mathrm{LogDIN}^{**}$	0.14	445	68	1086
11		$0.12^{*} + 0.33\mathrm{LogTP}^{***} + 0.23\mathrm{LogWT}^{***} - 0.07\mathrm{LogDIN}^{**} + 0.11\mathrm{LogTN}^{**}$	0.15	439	48	1085

Note: AIC, Akaike Information Criterion; df, degree of freedom.The significance of the regression coefficients is indicated by.*** $P < 0.001$. ** $0.001 < P < 0.01$. * $0.01 < P < 0.05$.

Table 3. Sensitivities of MCs hazard to different species-specific biomass, adopting the four-node networks.

	MCs hazard sensitivities								
Network	**WT**	**TN**	**SI**	**B_{MA}**	**B_{MW}**	**B_{MF}**	**B_{MI}**	**B_{MV}**	**B_{MN}**
A	1.80	1.52	4.49	**7.73**					
B	**2.66**	1.09	1.02		1.25				
C	**3.42**	2.19	2.53			0.59			
D	2.72	1.87	**4.96**				1.05		
E	**3.50**	1.74	0.58					2.60	
F	**3.92**	0.67	1.07						0.48

Table 4. Comparison of model performance and complexity, respectively based on Cohen's Kappa (K) and correctly classified instances (CCI), number of directed acyclic (Arcs), and conditional probabilities (CPs).

Network	CCI	K	Arcs	CPs
BN-B_M	0.81 ($\sigma = 0.04$)	0.29 ($\sigma = 0.04$)	5	496
BN-B_{MA}	0.94 ($\sigma = 0.02$)	0.55 ($\sigma = 0.05$)	4	170
BN-MCs	0.62 ($\sigma = 0.16$)	0.14 ($\sigma = 0.05$)	3	60

Note: See Fig. 4 for network definitions. BN-B_M, BN-B_{MA}, and BN-MCs refer to Bayesian network in Fig. 4a–c, respectively.

Table 5. Comparison of eight models performance in prediction of MCs with the same discretized data, based on correctly classified instances (CCI), Cohen's Kappa (K), root mean square error (RMSE), and relative absolute error (RAE).

Model	CCI	K	RMSE	RAE
BN[a]	0.759 (σ = 0.094)	0.403 (σ = 0.235)	0.394 (σ = 0.066)	0.681 (σ = 0.131)
BN	0.832 (σ = 0.079)	0.598 (σ = 0.188)	0.356 (σ = 0.080)	0.530 (σ = 0.143)
NN[b]	0.786 (σ = 0.085)	0.489 (σ = 0.200)	0.409 (σ = 0.090)	0.515 (σ = 0.180)
2-NN [d]	0.796 (σ = 0.084)	0.526 (σ = 0.189)	0.378 (σ = 0.073)	0.567 (σ = 0.139)
3-NN[e]	0.812 (σ = 0.078)	0.539 (σ = 0.192)	0.369 (σ = 0.066)	0.583 (σ = 0.123)
ID3	0.770 (σ = 0.090)	0.499 (σ = 0.186)	0.450 (σ = 0.098)	0.547 (σ = 0.214)
LR	0.827 (σ = 0.079)	0.576 (σ = 0.193)	0.360 (σ = 0.058)	0.659 (σ = 0.128)
SVM	0.820 (σ = 0.078)	0.552 (σ = 0.195)	0.413 (σ = 0.096)	0.440 (σ = 0.191)
MLP	0.795 (σ = 0.082)	0.496 (σ = 0.199)	0.406 (σ = 0.086)	0.509 (σ = 0.167)
C4.5	0.810 (σ = 0.079)	0.545 (σ = 0.188)	0.374 (σ = 0.068)	0.624 (σ = 0.120)
RF	0.816 (σ = 0.082)	0.549 (σ = 0.195)	0.360 (σ = 0.062)	0.611 (σ = 0.118)

Note: [a] BN excludes *Microcystis aeruginosa* and M. viridis as input variables.
[b˜d] k-Nearest Neighbor, parameter of k sets 1, 2, and 3 respectively.

Table 6. Probability table for total *Microcystis* biomass, *Microcystis aeruginosa* biomass hazard classes in response to different combinations of water temperature (WT), total phosphorus (TP) and dissolved inorganic nitrogen (DIN).

Conditions			**Probability of *Microcystis* class (%)**			**Probability of toxic *Microcystis* class (%)**		
WT	**TP**	**DIN**	**Low**	**Moderate**	**High**	**Low**	**Moderate**	**High**
<20			52.1	33.5	14.4	73.8	13.4	12.8
20-24			25.8	33.1	41.1	56.6	20.3	23.1
≥24			18.2	43.7	38.2	36.7	29.8	33.6
	<0.10		51.3	34.2	14.5	74.1	16.6	9.31
	0.10-0.20		42.3	36.6	21.1	68.7	14.5	16.9
	≥0.20		29.8	32.5	37.7	49.7	23.1	27.2
<20	<0.10		68.7	25.6	5.71	83.8	11.2	5.06
<20	0.10-0.20		52.3	33.9	13.8	78.6	8.68	12.8
<20	≥0.20		42.4	37.5	20.1	60.7	21.7	17.6
20-24	<0.10		31.3	42.8	25.9	64.6	22.4	13.0
20-24	0.10-0.20		33.9	36.1	30.0	64.0	19.6	16.4
20-24	≥0.20		11.8	23.5	64.7	40.7	20.1	39.2
≥24	<0.10		21.7	51.7	26.6	55.0	26.1	18.8
≥24	0.10-0.20		18.6	49.4	32.1	37.9	27.3	34.8
≥24	≥0.20		15.6	31.4	53.0	24.0	35.5	40.4
≥24	≥0.20	<0.4	33.3	33.3	33.3	16.0	16.0	**67.9**
≥24	≥0.20	0.4-1.5	10.7	33.5	55.8	35.3	37.6	27.2
≥24	≥0.20	≥1.5	8.44	22.5	**69.0**	0	53.1	46.9

Table 7. Probability of *M. aeruginosa* biomass and microcystin hazard classes for (1) initial conditions ; (2) simulated WT warming (WT ≥ 24 ℃); (3) simulated TN decrease (TN < 2.0 mg/L); (4) simulated both WT warming and TN decrease.

Hazard Class		**1)Initial conditions**	**2) WT warming**	**3) TN decrease**	**4) WT warming and TN decrease**
Toxic *Microcystis*class (%)	Low	63.6	36.7	60.8	30.6
	Mediate	17.7	29.8	21.6	30.9
	High	18.7	33.6	17.6	38.5
Microcystin class (%)	Low	29.1	13.7	32.1	17.9
	Mediate	43.3	45.9	37.3	30.6
	High	27.6	40.4	30.6	51.5

Note: Initial conditions refer to Bayesian networks in Fig. 4b and c.