Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems

Anna Rigosi,[1,11] Paul Hanson,[2] David P. Hamilton,[3] Matthew Hipsey,[4] James A. Rusak,[5] Julie Bois,[1] Karin Sparber,[6] Ingrid Chorus,[7] Andrew J. Watkinson,[8] Boqiang Qin,[9] Bomchul Kim,[10] and Justin D. Brookes[1]

[1]*Water Research Centre, University of Adelaide, Benham Building, South Australia 5005 Australia*
[2]*University of Wisconsin-Madison Center for Limnology, 680 North Park Street, Madison, Wisconsin 53706 USA*
[3]*Environmetal Research Institute, University of Waikato, Private Bag 3105, Hamilton 3240 New Zealand*
[4]*Aquatic Ecodynamics, University of Western Australia, 35 Stirling Highway, Crawley,*
Western Australia 6009 Australia
[5]*Dorset Environmental Science Centre, Ontario Ministry of the Environment and Climate Change, 1026 Bellwood Acres Road,*
Dorset, Ontario, Canada
[6]*Environmental Agency of the Autonomous Province of Bolzano/Bozen (APPA), Department Protection of Waterbodies,*
35 Amba Alagi, 39100 Bolzano, Italy
[7]*Federal Environment Agency, Corrensplatz 1, 14197, Berlin, Germany*
[8]*Seqwater, 117 Brisbane Street, Ipswich, Queensland 4305 Australia*
[9]*State Key Laboratory of Lake Science and Environment, Institute of Geography and Limnology, Chinese Academy of Sciences,*
Nanjing, China
[10]*Kangwon National University, Chuncheon, Gangwon-do, Republic of Korea*

Abstract. A Bayesian network model was developed to assess the combined influence of nutrient conditions and climate on the occurrence of cyanobacterial blooms within lakes of diverse hydrology and nutrient supply. Physicochemical, biological, and meteorological observations were collated from 20 lakes located at different latitudes and characterized by a range of sizes and trophic states. Using these data, we built a Bayesian network to (1) analyze the sensitivity of cyanobacterial bloom development to different environmental factors and (2) determine the probability that cyanobacterial blooms would occur. Blooms were classified in three categories of hazard (low, moderate, and high) based on cell abundances. The most important factors determining cyanobacterial bloom occurrence were water temperature, nutrient availability, and the ratio of mixing depth to euphotic depth. The probability of cyanobacterial blooms was evaluated under different combinations of total phosphorus and water temperature. The Bayesian network was then applied to quantify the probability of blooms under a future climate warming scenario. The probability of the "high hazardous" category of cyanobacterial blooms increased 5% in response to either an increase in water temperature of 0.8°C (initial water temperature above 24°C) or an increase in total phosphorus from 0.01 mg/L to 0.02 mg/L. Mesotrophic lakes were particularly vulnerable to warming. Reducing nutrient concentrations counteracts the increased cyanobacterial risk associated with higher temperatures.

Key words: Bayesian network; climate change; cyanobacterial blooms; multiple systems; nutrients; risk assessment; uncertainty.

INTRODUCTION

Cyanobacteria present a health risk through the production of toxins that degrade ecosystem services including water supply for irrigation, consumption, and recreation (Carpenter et al. 2011). These impacts and the need for removal of toxins and taste and odor-causing compounds from cyanobacteria in drinking water have high economic costs (Dodds et al. 2009). The geographical distribution of some bloom-forming cyanobacteria is increasing (Fristachi et al. 2009, Winter et al. 2011) and species historically observed in subtropical systems

(e.g., *Cylindrospermopsis raciborskii*) have recently invaded mid-latitude regions (Ryan et al. 2003, Briand et al. 2004, Sinha et al. 2012) and become more widely distributed. These changes have been attributed to an increase in water temperature and degree of stratification, to adaptation of phytoplankton, to new environmental conditions, or to nutrient enrichment leading to eutrophication (Paerl and Huisman 2008, Conley et al. 2009). However, it remains unclear how individual or cumulative impacts of changes in temperature, nutrients, global connectivity or other factors affect the growth of cyanobacteria (Hallegraeff 1993, Jöhnk et al. 2008, Brookes and Carey 2011, Huber et al. 2012).

Teasing out the relative importance of these drivers at a global scale and estimating the probability of cyanobacteria occurrence under different environmental

conditions using a simplified statistical framework would be extremely useful to enhance our understanding of bloom forming processes and support water management decisions to control cyanobacterial blooms under warmer conditions. Several works analyzed the statistical relationship between nutrient availability and cyanobacterial incidence including research on a large variety of European lakes (Carvalho et al. 2011, 2013, Dolman et al. 2012), although these studies have not explored the interaction with changing temperatures. This problem has been more frequently assessed at a site-specific scale (Carvalho and Kirika 2003, Arhonditsis et al. 2007, Elliott and May 2008), but not at a global scale due to challenges related to a lack of data covering broad gradients of nutrients and temperature as well as a lack of a comprehensive analytical framework. Covering a broad range of conditions will allow a straight forward application of the statistical model to different systems where information is too limited to develop other detailed models as deterministic or process-based ones. Additionally, calibrating models with data from lakes that span a broad latitudinal range and cover a broad range of nutrient conditions should result in models that are more widely applicable than those calibrated to a single lake.

Bayesian networks represent a useful framework to address the latter calibration problem because they can integrate multiple sources of information to estimate model parameter values and because they account for result uncertainty, thereby avoiding reliance on a single deterministic outcome that does not reflect the inherent natural ecosystem variability (Arhonditsis et al. 2007). Bayesian networks are graphically based and so are easier to understand than many modelling tools, they are a powerful communication instrument representing uncertainties and they can be easily updated as data or knowledge become available.

A Bayesian network (BN) consists of a combination of graphical links and statistical correlations among the most important variables in the studied system. Variables are represented as nodes and unidirectional dependence relationships are developed between one variable and another. BNs use probabilistic rather than deterministic expressions to describe the relationship among variables. Each arrow that links a "parent" node to a "child" node represents a conditional probability distribution that describes the likelihood of each value of the child node given the combination of values of the parent nodes (Borsuk et al. 2004, Hamilton et al. 2007). BNs use the network structure to calculate the probability that certain events will occur and how these probabilities will change given subsequent observations or management interventions. Bayesian models provide a useful framework for analyzing alternative scenarios of system change (Bromley et al. 2005, Quinn et al. 2013).

Bayesian statistical inference has been shown to assist in parameter estimation and hypothesis testing, especially in the field of environmental decision-making (Ellison 1996). Moreover, it has been recognized as a useful framework for ecological modelling and resource management, particularly in representing population variability and supporting decision-making processes (Reckhow 1999, Bromley et al. 2005, McCann et al. 2006, Quinn et al. 2013). Further details on Bayesian statistical inference for ecological research are presented by Haenni et al. (2011) and Ellison (1996).

Few examples exist of the application of BNs to study eutrophication and, in particular, cyanobacterial bloom development. To predict site-specific algal blooms statistical models such as neural networks have been used extensively (Maier et al. 1998, Lee et al. 2003, Muttil and Chau 2006) or, in other cases, Bayesian statistical knowledge has been combined with deterministic models (Arhonditsis et al. 2006, 2007). Two examples of where Bayesian analysis was applied to cyanobacterial bloom development are for *Lyngbya* sp. in Deception Bay and Moreton Bay, Queensland, Australia (Hamilton et al. 2005, 2007). These examples showed the utility of BNs to incorporate different sources of information and to analyze the risk of cyanobacterial bloom occurrence. The most influential factors for bloom development in these systems were identified to be water temperature, nutrient, and light availability (Hamilton et al. 2007). However, while BN models are useful at a particular site to better understand localized impacts of climate change and eutrophication, there is a need for models that can be conditioned on ecosystem data from a wide range of climatic and geomorphic contexts to apply to lakes where only limited data is available. Furthermore a model developed with a broad range of temperature and nutrient inputs will enable extrapolation through time to predict cyanobacterial risk with temperature or nutrient changes. These estimations will provide a general understanding of how changes in temperature and in nutrient availability interact and will allow estimating how much nutrients should be reduced to counteract the effects of warming.

Our study utilizes data from 20 lakes located at different latitudes to address this gap and demonstrates the application and utility of Bayesian models for cyanobacterial bloom predictions across a broad geographic range. The main objectives of this study were to (1) determine which environmental factors were the most important explaining variation in the abundance of cyanobacteria observed in multiple lakes and reservoirs; (2) analyze the significance of interactions among contributing environmental factors; and (3) test BNs to estimate how changes in environmental conditions may alter the probability of cyanobacterial bloom development, in order to support informed water management decisions. To achieve these aims, BNs structures were built starting from a conceptual model of the possible drivers controlling cyanobacterial blooms. Then, they were populated using the broad

ANNA RIGOSI ET AL.

TABLE 1. Characteristics of the 20 lakes included in the database for the development of the statistical model.

Ref no.	Lake name	Location	Latitude	Area (km²)	MD (m)	KCC†	Trophic state	Data‡
1	Myponga	South Australia, AU	35°24′13″ S	2.5	43.9	Csb	ME	2007–2012 (216)
2	Annie	Florida, USA	27°12′25″ N	0.34	20	Cfa	O	2008–2009 (21)
3	Meiliang (Taihu)	Jiangsu, China	31°32′58″ N	2338	2.6	Cfa	E	2008–2010 (33)
4	Soyang	South Korea	32°32′24″ N	63.9	85	Dwa	O	2010–2011 (33)
5	Mendota	Wisconsin, USA	47°59′24″ N	39.4	25.3	Dfa	E	1995–2008 (91)
6	Monona	Wisconsin, USA	43°4′9″ N	13.3	22.6	Dfa	E	1995–2009 (4)
7	Erken	Sweden	59°51′ N	23.7	21	Dfb	M	2010 (19)
8	Feeagh	Ireland	53°56′50″ N	3.9	45	Cfb	O	2008–2010 (15)
9	Tegel	Germany	52°35′0″ N	3.1	16	Dfb	E	1987–2006 (227)
10	Rotorua	New Zealand	38’04′16″ S	79.0	40	Cfb	E	2003–2011 (19)
11	Harp	Ontario, Canada	45°22′48″ N	0.7	37.5	Dfb	O	2011 (13)
12	Thomson	Victoria, AU	37°41′49″ S	22.8	165.7	Cfb	O	2007–2011 (40)
13	Upper Yarra	Victoria, AU	37°41′0″ S	5.5	81.35	Cfb	O	2005–2012 (63)
14	Yan Yean	Victoria, AU	37°59′53″ S	5.4	3	Cfb	E	2004–2012 (61)
15	Tarago	Victoria, AU	37°33′22″ S	3.4	21.7	Cfb	O	2004–2012 (75)
16	Advancetown (Hinze dam)	Queensland, AU	28°3′0″ S	207	43	Cfa	ME	1999–2011 (53)
17	Little Nerang	Queensland, AU	28°8′38″ S	35.2	30	Cfa	ME	1999–2011 (69)
18	Samsonvale (North Pine)	Queensland, AU	27°16′19″ S	348	35	Cfa	E	1997–2011 (271)
19	Somerset	Queensland, AU	27°6′50″ S	1340	35	Cfa	E	1997–2011 (147)
20	Wivenhoe	Queensland, AU	27°23′38″ S	7020	38	Cfa	E	1997–2011 (144)

Notes: Mean (with SD in parentheses) are shown for all the physical, chemical, biological variables included in the database. Variables are MD, maximum depth; KCC, Köppen climate classification; PAR, photosynthetically active radiation; WS, wind speed; WT, water temperature; z_{mix} : z_{eu}, ratio between mixing depth and euphotic depth; TN, total nitrogen, TP, total phosphorus; CyanoHazard, cyanobacterial bloom hazard, based on cyanobacterial abundance. Abbreviations are ME, meso-eutrophic; O, oligotrophic; E, eutrophic; M, mesotrophic; AirT, air temperature; AU, Australia; na, not available.

† Köppen climate classification: Csb, temperate/mesothermal (dry summer, subtropical or mediterranean climate); Cfa, temperate/mesothermal (humid subtropical climate); Dwa, continental/microthermal (hot summer, continental climate); Dfa (Dwa), continental/microthermal (hot summer, continental climate); Dfb, continental/microthermal (warm summer continental or hemiboreal); Cfb, temperate/mesothermal (oceanic climate); Dfb, continental/microthermal (warm summer continental or hemiboreal).

‡ Not all the years indicated are complete. Numbers in parentheses indicate the number of complete cases available for each lake when considering three variables: cyanobacteria abundance, water temperature, and TP.

range of environmental conditions found in 20 lakes, in order to identify the most important drivers. A previous analysis of the 20-lake database, also included here, was essential to identify the range of variation of the environmental factors and check for correlations among predictor variables.

METHODS

The limnological data collected were from 20 lakes (see Plate 1) with areas ranging from 0.7 to 7010 km². They spanned broad gradients in latitude, lake morphometry, nutrient concentrations, water temperature, and degree of water column mixing (Table 1). With this data, we first determined the range of variability in each property as well as the most important drivers of cyanobacterial blooms and the covariation among these drivers. We then built a Bayesian network model that predicted the probability of cyanobacterial blooms including uncertainty in model output.

Lake database assemblage

A database was compiled from data available within the Global Lake Ecological Observatory Network (GLEON) and from databases provided by other collaborators (see Acknowledgments). Table 1 provides a list of the lakes, a summary of their characteristics and the periods of data availability. The following variables

were included: cyanobacterial abundance (cells/mL), total phosphorus (TP, mg/L), total nitrogen (TN, mg/L), water temperature at surface (°C), air temperature (°C), photosynthetically active radiation (PAR, W/m²), wind speed (m/s), mixing depth (m), euphotic depth (m), ratio between mixing and euphotic depth, maximum lake depth (m), and latitude (degrees). The mixing depth (z_{mix}) was defined as the depth of the surface mixed layer, the portion of the water column influenced by wind and convective cooling where both temperature and density are vertically homogeneous (Read et al. 2011). In this work, given that not all lakes had continuous water column temperature profiles and mostly discrete samples were available (e.g., every 1 m), the mixing depth was calculated as the layer at which the vertical temperature gradient was <0.2°C/m. The first sampling depth at which this gradient was exceeded was considered to be the mixing depth. Temperature differences were calculated every 1 m starting from 1 m depth, in order to determine the seasonal thermocline location (rather than the diurnal thermocline). The euphotic depth (z_{eu}) was defined as the depth at which the light intensity was 1% of that immediately below the water surface (Grobbelaar and Stegmann 1976) and was calculated from Secchi depth measurements as $z_{eu} = a(\text{Secchi depth})^b$ where $a = 4.1865$ and $b = 0.73$, following (Martin and McCutcheon 1999). Chemical

TABLE 1. Extended.

AirT (°C)	PAR (W/m^2)	WS (m/s)	WT (°C)	z_{mix}:z_{eu}	TN (mg/L)	TP (mg/L)	CyanoHazard (cells/mL)
15.8 (5.1)	94.1 (46.8)	3.24 (1.4)	17.4 (3.7)	1.65 (1.07)	1.06 (0.15)	0.04 (0.03)	7297 (23633)
22.4 (4.4)	94.8 (21.51)	1.62 (0.41)	25.65 (4.6)	0.11 (0.11)	0.29 (0.06)	0.006 (0.0015)	0.47 (0.38)
16.43 (9.35)	181.5 (181.13)	3.5 (0.66)	16.84 (9.8)	0.76 (0.35)	3.06 (1.21)	0.115 (0.063)	58483 (88147)
13.4 (10.5)	258.7 (86.6)	1.08 (0.32)	16.7 (8.5)	3.49 (3.62)	1.43 (0.15)	0.0086 (0.0025)	1471 (4720)
12.5 (9.6)	86.38 (31.6)	3.9 (1.08)	15.7 (7.8)	0.89 (0.89)	1.04 (0.34)	0.081 (0.05)	140263 (241760)
11.3 (9.7)	83.2 (32.8)	3.4 (0.8)	16.8 (8.1)	1.11 (1.06)	1.14 (0.2)	0.09 (0.02)	201174 (313043)
12.7 (9.9)	84 (41.5)	3.6 (0.65)	17.1 (4.1)	0.70 (0.48)	0.68 (0.066)	0.027 (0.009)	5677 (7496)
9.8 (3.7)	17.8 (12.2)	5.0 (1.4)	10.7 (4.0)	4.8 (2.7)	0.20 (0.12)	0.008 (0.002)	68 (178)
na	na	na	14.2 (6.5)	1.3 (0.8)	5.2 (2.5)	0.056 (0.037)	11701 (31478)
12.6 (3.6)	280.5 (126)	3.3 (0.9)	15.9 (4.2)	2.0 (1)	0.47 (0.25)	0.032 (0.013)	5541 (8926)
10.9 (6.3)	73.4 (38.6)	1.5 (0.6)	13.5 (6.7)	0.33 (0.2)	0.23 (0.03)	0.005 (0.001)	0.16 (0.14)
13.2 (4.5)	153.8 (56.1)	4.1 (0.6)	15.1 (4.81)	na	0.309 (0.069)	0.0045 (0.0029)	2671 (5212)
12.3 (4.05)	123.6 (52.5)	2.78 (0.38)	15.9 (4.84)	na	0.2656 (0.0615)	0.0075 (0.0080)	2198 (3836)
15.15 (3.86)	138.85 (52.01)	3.38 (0.899)	17.116 (4.50)	na	0.703 (0.44)	0.024 (0.0246)	14870 (88732)
14.55 (3.67)	108.26 (42.28)	1.53 (0.88)	17.97 (4.58)	na	0.6064 (0.179)	0.0136 (0.0071)	4385 (7134)
na	na	na	22.4 (3.9)	1.1 (0.7)	0.33 (0.1)	0.019 (0.015)	12447 (35760)
na	na	na	21.1 (3.9)	0.7 (1)	0.30 (0.10)	0.021 (0.011)	8935 (29347)
na	na	na	22.9 (3.8)	3.5 (2.3)	0.55 (0.1)	0.019 (0.09)	39380 (80569)
na	na	na	23.0 (4.3)	2.8 (2.9)	0.58 (0.18)	0.030 (0.036)	124614 (155407)
21 (3.8)	79 (24)	2.9 (0.6)	22.9 (3.9)	4.0 (4.1)	0.50 (0.10)	0.023 (0.019)	109611 (107471)

and biological variables were generally measured less frequently than physical variables: cyanobacteria abundance was available monthly, or weekly (in 19 lakes), or twice weekly (Myponga Reservoir). Samples represented depth integrated epilimnetic concentrations.

To account for the fact that conditions preceding bloom development (e.g., resource depletion, turbulence, fluctuating light) strongly influence bloom occurrence (Reynolds 2006, Hamilton et al. 2009), a one-week time lag was selected between biological and environmental variables for data aggregation. The choice of this time lag is justified by the fact that in most of the lakes considered, cyanobacterial bloom development occurred over a few days and that several authors have recognized that ∼5–15 days is a time scale on which a phytoplankton group may become dominant (Reynolds et al. 1993, Maier et al. 1998, Gallina et al. 2011). Average values of meteorological variables (air temperature, radiation, and wind speed) were calculated for seven days prior to the sampling of cyanobacterial abundance. Water chemistry data were available on the same day as cyanobacteria were sampled. For water column physical variables (e.g., mixing depth and surface water temperature) a seven-day average was calculated when water temperature data were monitored remotely, otherwise a single value was considered when cyanobacteria were sampled seven days beforehand.

Soluble reactive phosphorus, total nitrogen, and nitrate concentrations were available only for a few of the lakes, so total phosphorus was selected as the nutrient determining cyanobacterial abundance. Using TP is in line with phosphorus being the limiting nutrient in the majority of situations; furthermore it reflects capacity to support phytoplankton biomass to a greater degree than soluble reactive phosphorus (Hudson et al. 2000). No consideration was given to the ratio of TN:TP

as this has been shown to be more influential on cyanobacterial species composition rather than total cyanobacterial biomass (Fujimoto et al. 1997, Levine and Schindler 1999, Downing et al. 2001, Nalewajko and Murphy 2001).

Lake database analysis

A multivariate analysis was used to identify the most important variables explaining ecological differences between lakes. Lakes were grouped depending on their similarities and the main explanatory variables were identified. First, three variables, available for all the lakes and for all sampling days, were considered: surface water temperature, total phosphorus concentration, and cyanobacterial abundance. The same analysis was then repeated with three additional variables: mixing depth, latitude, and maximum lake depth. Averages of the variables during the summer period were used (October–April for the Southern Hemisphere and May–September for the Northern). Phytoplankton data were standardized by dividing observed cyanobacteria density by the maximum cyanobacteria density over the summer period. As recommended for abundance and biomass data, a dissimilarity matrix was first built on the basis of the Bray-Curtis dissimilarity measures (Clarke and Gorley 2006). Then, a redundancy analysis (RDA) was applied, which is a constrained ordination where the axes are forced to be linear combinations of the variables (Clarke 1993). As an outcome of the RDA, lakes were spatially grouped, and the main explanatory variables were determined visually and statistically. The analysis was conducted with the software Primer-E (Clarke and Gorley 2006). Relationships between phosphorus, temperature, and cyanobacteria in the database were further analyzed generating histograms.

PLATE 1.   Meiliang Bay, Lake Taihu, Jiangsu, China, one of the 20 lakes included in our study. Photo credit: Jianrong Ma.

Bayesian network

Model structure.—Guidelines were followed in the creation and testing of the Bayesian ecological network (Marcot et al. 2006, Pollino and Henderson 2010). First, the objective of the model and its final node (cyanobacterial bloom development) were defined; second, a conceptual model of possible drivers was generated based on knowledge from the literature and on expert knowledge; and third, model nodes and states were established.

The probabilities of the marginal nodes and the conditional probabilities between nodes (conditional probability tables, CPTs) were populated using the empirical data available from the 20 lakes. Once the network structures were populated with data, their prediction ability was tested for ability to predict cyanobacterial abundance, followed by revision of the states and variables to improve model fit. The BN was constructed using the software Netica (Norsys Software Corporation, Vancouver, British Columbia, Canada). Details of the algorithms used by Netica to exact general probabilistic interference can be found in Spiegelhalter et al. (1993) and Jensen (1996). The model application focused on the probability of cyanobacterial bloom development as an endpoint. Hazard levels for cyanobacterial blooms were defined as low, when cyanobacteria abundance was $\leq 0.2 \times 10^5$ cells/mL; moderate, between 0.2×10^5 and 1×10^5 cells/mL; and high, when $>1 \times 10^5$ cells/mL (Chorus and Bartram 1999). These thresholds were used to define the three states of the network endpoint node called "cyanobacterial hazard." Three BNs of differing complexity were generated using

three, four, and nine nodes. The BNs with different number of nodes were developed to identify the most important drivers affecting cyanobacterial hazard and to evaluate networks ability to predict an alteration in bloom development. The number of cases to generate each BN, depending on data availability, was monitored and specified. Thresholds and corresponding number of states used for each node, both affecting the model results, were established based on literature (e.g., total phosphorous concentration thresholds for oligotrophic, mesotrophic, and eutrophic systems), analysis of data distribution, and knowledge achieved by running the model and testing its predictive ability using different states and different threshold values (Appendix).

Model evaluation and sensitivity analysis.—To assess the predictive accuracy of different BNs, the network was first generated using a subset of observed data corresponding to 80% of the observations in the database for each lake. The network was then tested to evaluate how well the diagnosis matched observations using the remaining 20% of observations. Due to the fact that the network results are dependent on the cases used to build it, the procedure was repeated three times, randomly selecting data from the complete database. As a result of the evaluation a percentage error rate was determined which is the fraction of misclassified cases as a proportion of the classifications made.

To determine which parts of the model most affected the variable of interest (i.e., cyanobacterial bloom hazard) a sensitivity analysis was undertaken on the endpoint node. The sensitivity analysis generated nodal values (percentages) that were compared between nodes
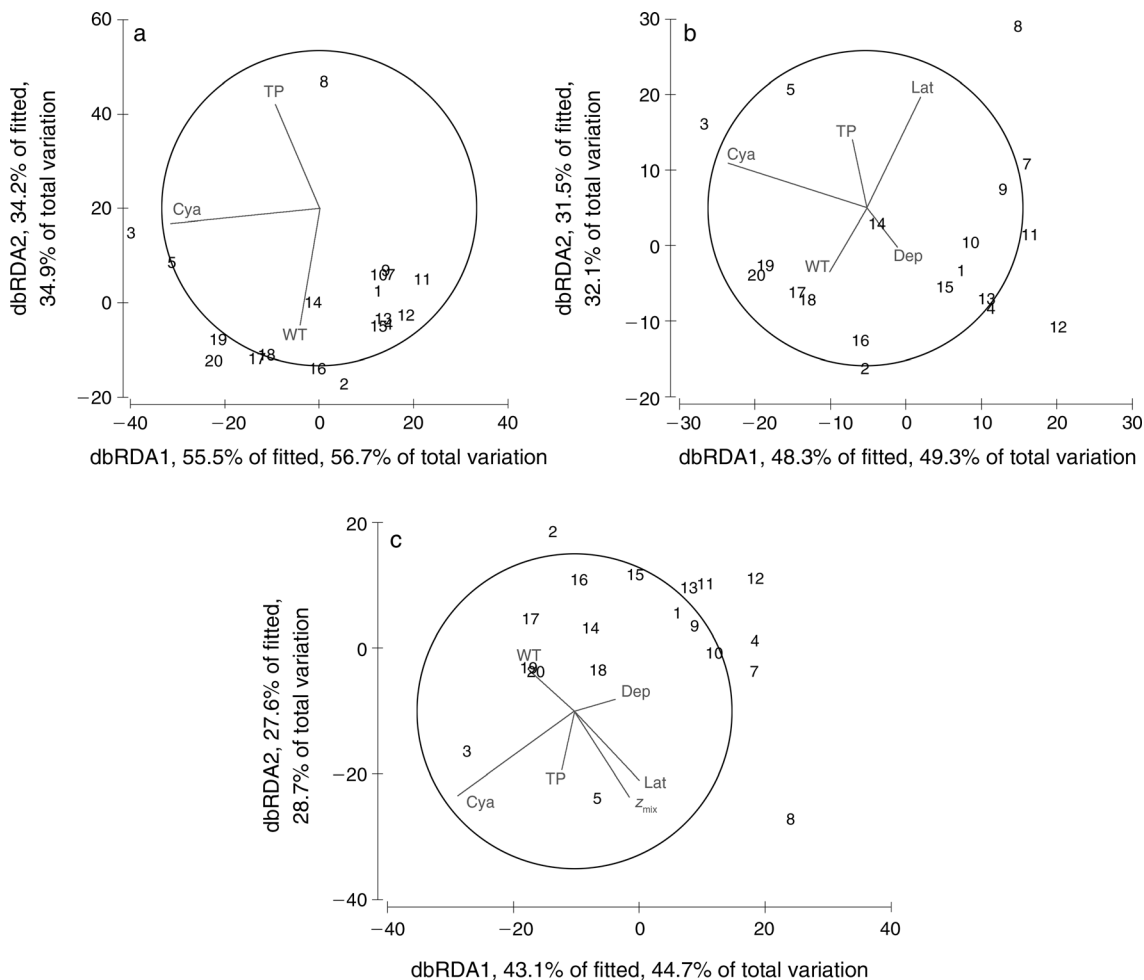
FIG. 1. Distance-based redundancy analysis (dbRDA) including (a) three variables, (b) five variables, and (c) six variables. Each number represents one lake, as listed in Table 1; axes in the circle represent the different explanatory variables: Cya (cyanobacterial abundance), WT (water temperature), TP (total phosphorus), Dep (maximum depth), Lat (latitude), z_{mix} (mixing depth).

of the same network to establish which variable had the greatest effect on the assigned bloom hazard value.

Scenarios.—After the model was evaluated, future potential environmental scenarios were developed as input to the BN. The scenarios included a trend of warming of surface water temperature by 0.037°C/yr (Schneider and Hook 2010), combined with different ranges of phosphorus concentration (TP; e.g., low, TP < 0.02 mg/L; medium, TP = 0.02–0.10 mg/L; high, TP ≥0.1 mg/L). Changes in the probability of cyanobacterial bloom hazard (low, moderate, and high) were calculated for each of the scenarios.

RESULTS

Lake database analysis

The dominant environmental variables that distinguished cyanobacterial abundance between lakes were assessed by RDA (Fig. 1). Lakes 3, 5, 19, and 20 (refer to Table 1) were characterized by high average

cyanobacteria abundance and were aligned with the cyanobacteria axes (Fig. 1a). Lake 8 was characterized by lower levels of total phosphorus with respect to the other lakes; while lakes 2, 16, 17, 18, 19, and 20 were characterized by high water temperatures. Adopting three variables (Fig. 1a) yielded a higher proportion of the fitted variation for both axes than including additional variables, such as mixing depth, latitude, and maximum depth (Fig. 1b, c). The fact that including more variables did not improve the fit (dropping from 89.7% to 79.8%) suggests that water temperature and total phosphorus, together with cyanobacterial abundance, were the most important explanatory variables describing variability between lakes. Latitude influenced how lakes grouped together (Fig. 1b) but neither latitude nor depth were more strongly related to cyanobacterial abundance than total phosphorus and temperature. Fig. 1c shows the importance of z_{mix} compared with the other variables analyzed. However, water temperature

ANNA RIGOSI ET AL.

and total phosphorus were still the best descriptors of cyanobacterial abundance (fit decreased to 71.8%).

Histograms were generated for the percentage of cyanobacteria abundance corresponding to specific hazard classes (low, moderate, and high; Chorus and Bartram 1999) vs. total phosphorus and surface water temperature (Fig. 2). The cases characterized by low, moderate, and high hazard were, respectively, 52%, 28%, and 20% in the 20 lakes. Fig. 2c shows, for example, that at low total phosphorous (TP) concentrations the cases of high hazard were more frequent if water temperatures were high (>23°C). Low cyanobacterial abundances (classified in the low hazard category) occurred more frequently at lower temperatures (<21°C) (Fig. 2a). Graphical results obtained representing the current database suggest that the effect on cyanobacterial abundance of interactions between nutrients and water temperature (WT) is not additive, although few cases with high TP were available in the data set. It is notable that there appears to be a high dependence of cyanobacterial abundance on WT at lower TP concentrations.

The number of blooms classified as high hazard changed with different combinations of WT and TP in the lakes. Eighty-two per cent of the high hazard events occurred when WT was between 20°C and 30°C and about 60% of these occurred when TP was between 0.01 and 0.03 mg/L. This suggests that blooms classified as high hazard are much less likely to occur if phosphorus concentrations are low, i.e., TP < 0.01 mg/L. Moreover, when TP was low, blooms were most likely to occur when WT was high; however, as TP increases, cases of hazardous blooms were still observed in the data base at relatively low temperatures (WT < 15°C).

Bayesian network results

Model structure.—Three different network structures were adopted. A simplified network with three nodes was used to analyze the relationship of cyanobacteria abundance to total phosphorus and water temperature (Fig. 3a). A network of four nodes was adopted to analyze cyanobacteria sensitivity to additional environmental factors, including mixing depth, euphotic depth, meteorological conditions, or the depth and the location of the lake (Fig. 3b). The most complex network included nine nodes (Fig. 4): Cyanobacterial hazard (CyanoHazard), total phosphorus (TP), surface water temperature (WT), ratio between mixing depth and euphotic depth ($z_{mix}:z_{eu}$), photosynthetically active radiation (PAR), wind speed (WS), air temperature (AirT), latitude, and maximum lake depth (depth). Node definitions are given in Appendix: Table A2. The analysis of scenarios was conducted with the three-node network, including one additional state for each node (Fig. 3c). Repeated simulations showed that the probability distribution of the cyanobacteria hazard was affected by the network structure: number states and different thresholds were analyzed before proceeding to the sensitivity analysis (Appendix).

Model evaluation and sensitivity analysis.—Sensitivity analysis results conducted with the three-node net showed that, given the available data, the cyanobacterial hazard was more sensitive to WT than to TP; 20.3% and 0.12%, respectively. Using different case files (testing files including 80% of data) the maximum value found for sensitivity to TP was about 0.5%. To identify which other variables could be important in controlling cyanobacterial blooms, the sensitivity analysis was repeated, including a new parent node and adopting a four-node network (Fig. 3b). Variables that may directly affect the growth of cyanobacteria (z_{mix}, z_{eu}, PAR, $z_{mix}:z_{eu}$) were connected one-at-a-time to the end-point node. Sensitivity values defined as percentage hazard sensitivities are shown for each of the four-node networks in Table 2. It was observed that z_{mix}, z_{eu}, and $z_{mix}:z_{eu}$ were factors to which cyanobacteria abundance was more sensitive compared with PAR. Finally, in testing the nine-node network (Fig. 4), the error rates obtained varied between 17% and 22.6%. Cyanobacteria sensitivity to the other nodes in the network are listed in order of their importance: WT (14.5%), AirT (5.91%), $z_{mix}:z_{eu}$ (1.32%), TP (1.1%), depth (0.35%), latitude (0.23%), PAR (0.11%), WS (0.02%). Air temperature was also identified as an important factor but this was likely due to correlation with water temperature. Sensitivity values also indicated that lake depth and location were not as important in predictions of cyanobacterial abundance as TP and $z_{mix}:z_{eu}$. It should be noted that this complex network included a reduced number of observations (271 complete cases). Due to lack of meteorological data, only 10 lakes were included: 1, 2, 3, 4, 5, 7, 8, 10, 11, 20, and only a small number of cases were available for shallow lakes.

Scenarios.—The three-node network allowed quantification of the probability of low, moderate, and high cyanobacterial abundances given particular conditions of WT and TP. The error rate of the network ranged between 32% and 37% based on use of different test files. Thus, using WT and TP the probability of making a valid prediction of cyanobacterial hazard was about 60%.

The probability of high cyanobacteria abundance increased with increasing TP concentrations as well as increasing WT. When combining TP and WT, probabilities varied, demonstrating an interaction rather than an additive effect of these two factors (Table 3). At low WT the probability of high hazardous blooms was low and moderately hazardous abundances were more likely to occur at higher TP concentrations. At intermediate WT there was evidence of dependency on TP for high hazardous blooms, while, when temperatures were above 24°C, high hazardous blooms would occur even at low TP concentrations (Table 3). Moreover, at low and intermediate TP, high hazardous blooms were more likely to occur at higher WT (Table 3). The number of cases with high TP concentration (e.g., >0.05 mg/L) in
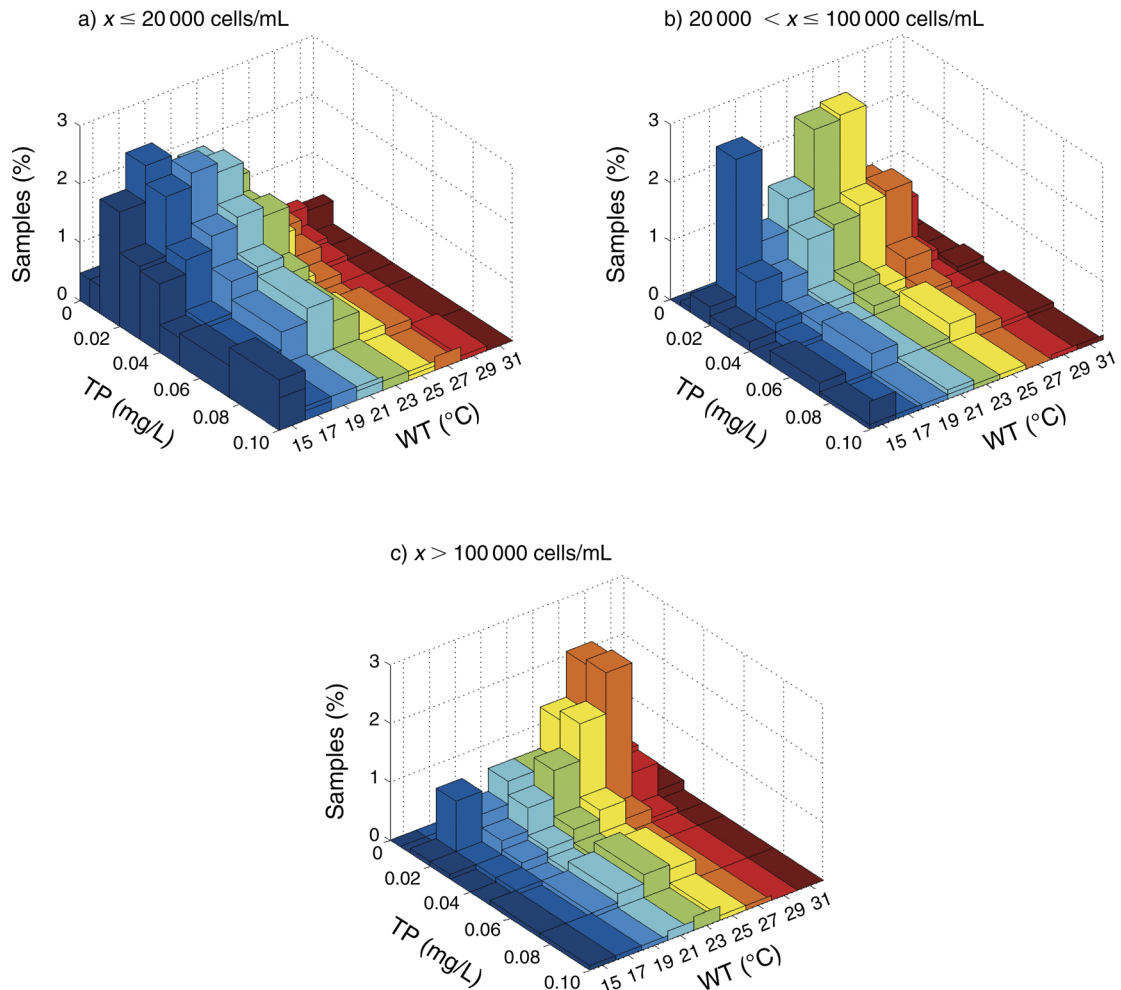
Fig. 2.   Percentage of samples (cases) from all the lakes classified as (a) low, (b) moderate, and (c) high hazardous blooms based on the cyanobacterial abundance (*x*) observed. Each bin represents a particular combination of the environmental conditions observed (total phosphorus and water temperature).

the 20-lake data base was insufficient to allow for a clear identification of trends of hazardous event occurrences corresponding to that condition.

To evaluate the effect of global warming trends, a three-node network with additional states (Fig. 3c) was employed. Increasing WT from state b (20–24°C) to state c (24–28°C) increased the cyanobacterial high hazardous bloom probability 22.6%. Modifying the TP from state b (0.01–0.02 mg/L) to state c (0.02–0.03 mg/L) increased high hazardous bloom probability about 4.6% (Table 4). Thus, a 5% increase in the probability of cyanobacterial high hazardous blooms was obtained either by increasing WT by 0.8°C or increasing TP by 0.01 mg/L. When using the network to test scenarios, we focused in particular on the change between oligotrophic to mesotrophic conditions (TP from 0.01 to 0.02 mg/L), but also included a state for eutrophic cases (TP > 0.03 mg/L; Fig. 3c). Changes in probabilities of moderate and low hazardous events are given in Table 4. The

changes in probability of cyanobacterial high hazardous blooms for eutrophic, mesotrophic, and oligotrophic conditions, when WT was increased by 4°C, were respectively: 13.9%, 27.1%, and 5%, showing the high vulnerability of mesotrophic systems to a change in temperature.

DISCUSSION

Predicting and managing cyanobacteria risk presents a major challenge for researchers and water resource managers. A comprehensive understanding of the causal factors leading to cyanobacterial blooms is lacking (Oliver et al. 2012), which limits the ability to predict cyanobacterial risk. Several different modelling approaches have been adopted to predict the magnitude and timing of cyanobacterial blooms and Rigosi et al. (2010) provide an extensive review of empirical and deterministic approaches that include key ecosystem variables and components of cyanobacterial physiology.
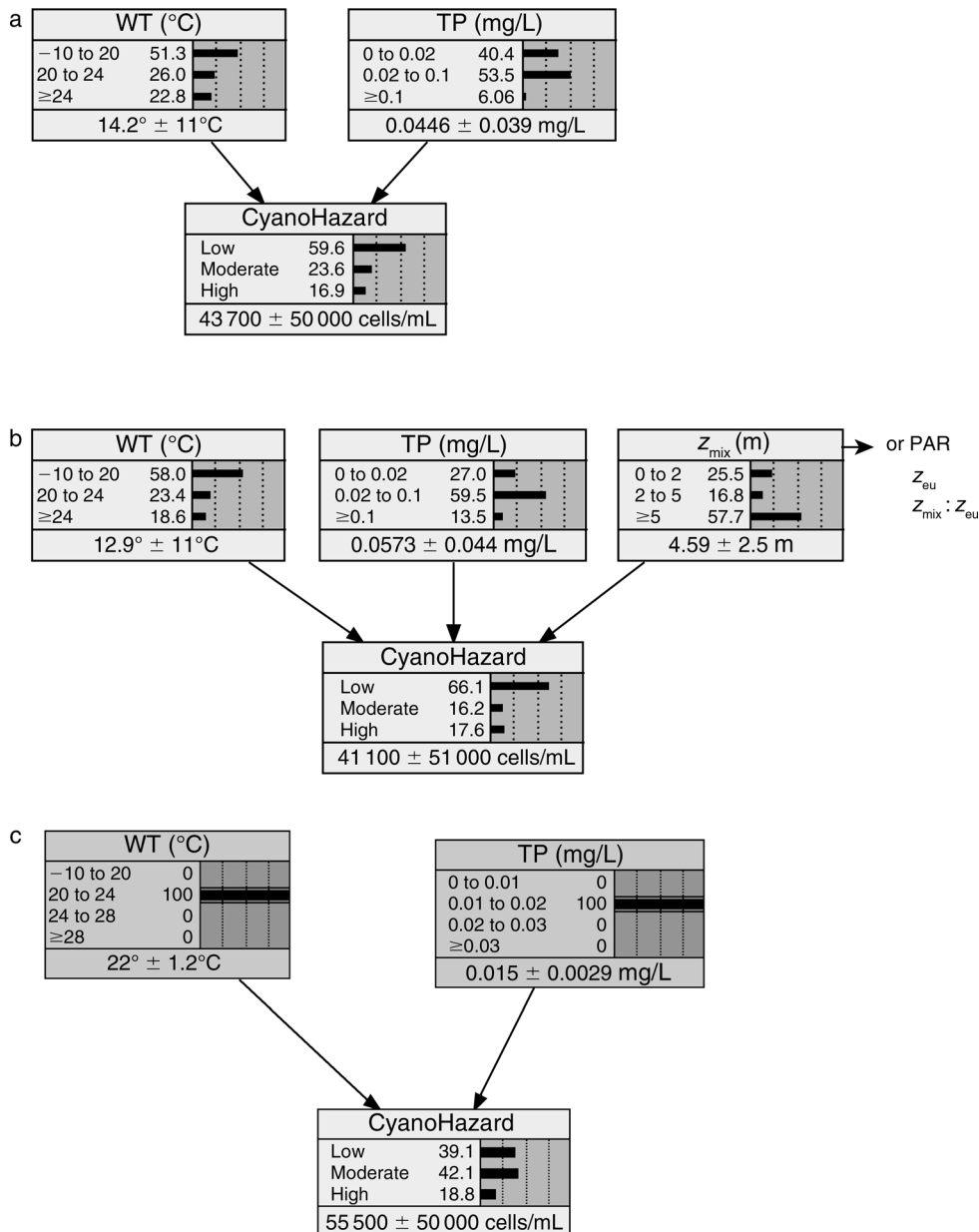
Fig. 3. Bayesian network structure for assessing cyanobacteria abundance (cells/mL) represented as CyanoHazard. (a) Most simplified network with only two parents (water temperature [WT] and total phosphorus [TP]); (b) one additional parent used in the sensitivity analysis (mixing depth [z_{mix}]); (c) network with four states used for testing scenarios; initial conditions ($20 < WT \leq 24$°C and $0.01 < TP \leq 0.02$ mg/L) are selected. Bars indicate probabilities (%) and values on the bottom of each node represent means and SD. In panel b, the arrow indicates that the z_{mix} node, when testing other network structures, was replaced alternatively with different nodes (photosynthetically active radiation [PAR], euphotic depth [z_{eu}], $z_{mix}:z_{eu}$).

In the present study we adopted a novel approach using a Bayesian network to identify casual factors for cyanobacterial blooms and cyanobacterial risk over a broad range of latitudes using a 20-lake database. The network provided an estimate of the probability of cyanobacteria occurring at particular magnitudes, corresponding to classes commonly used to define the level of risk (Chorus and Bartram 1999), using empirical relationships between cyanobacterial abundance and key environmental parameters.

The Bayesian modelling revealed that three factors contributed most to high cyanobacteria abundance (given as a probability): surface water temperature followed by total phosphorus and the ratio between mixing depth and euphotic depth. The variable $z_{mix}:z_{eu}$ is used to express cyanobacteria light exposure within
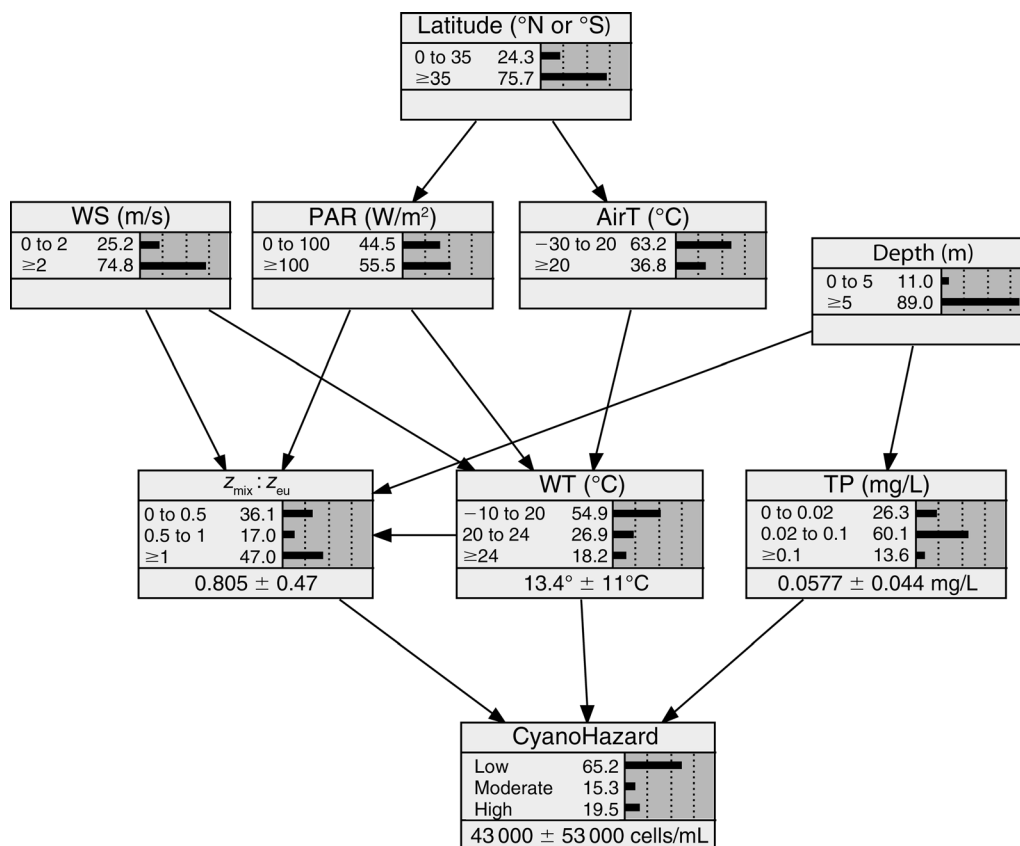
FIG. 4. Bayesian network structure including nine nodes: latitude, wind speed (WS), photosynthetically active radiation (PAR), air temperature (airT), maximum lake depth (Depth), ratio between mixing depth and euphotic depth (z_{mix}:z_{eu}), surface water temperature (WT), total phosphorus (TP) and cyanobacterial bloom hazard (CyanoHazard) based on cyanobacterial abundance (cells/mL).

the surface mixed layer. Similar factors were identified as dominant variables by Hamilton et al. (2007) when applying Bayesian networks in Deception Bay (Queensland) to assess the risk of *Lyngbya majuscula* blooms. In that case nutrients, water temperature, redox state of bottom sediments, current velocity, and light were the dominant variables. The statistically based artificial neural network used in a study of the Murray River (South Australia) identified water temperature and river flow as the predominant controls on the magnitude and duration of cyanobacteria growth (Maier et al. 1998).

Different cyanobacterial species have different light, temperature, and nutrient requirements and may display different physiological responses to these environmental variables (Reynolds 1997, Carey et al. 2012, Oliver et al. 2012). Furthermore the risk associated with different species may vary depending upon the type of toxin, or taste or odorous compounds produced. The factors generating hazardous blooms can be species and location dependent (Anderson et al. 2002), however, analyses that span across multiple lakes and latitudes offer insights into what trajectories may be observed with increases in temperature or nutrients. Kosten et al. (2012), using a one-year data set from 143 lakes in a latitudinal transect, showed that the relative cyanobacterial abundance in the community increased with

TABLE 2. Sensitivities of cyanobacteria hazard to factors, adopting the four-node networks.

Network	Parent Nodes	Cyanobacterial hazard sensitivities					
		WT	TP	z_{mix}:z_{eu}	z_{mix}	z_{eu}	PAR
A	WT, TP, z_{mix}:z_{eu}	12.6	0.753	**2.13**			
B	WT, TP, z_{mix}	13.7	1.54		**1.7**		
C	WT, TP, z_{eu}	11.9	0.41			**1.92**	
D	WT, TP, PAR	13.5	**3.06**				0.753

Notes: See Table 1 for factor definitions. Values in boldface type show the highest sensitivity values.

ANNA RIGOSI ET AL.

TABLE 3. Probability table for cyanobacterial bloom hazard classes in response to different water temperature and total phosphorus values.

Conditions		Probability of CyanoHazard class bloom (%)		
WT (°C)	TP (mg/L)	Low	Moderate	High
	<0.02	58.1	26.4	15.5
	0.02–0.10	61.5	21	17.5
	>0.10	52.4	26.8	20.8
<20		78.4	16.3	5.3
20–24		53.7	27.4	18.9
>24		23.7	35.5	40.8
<20	<0.02	75.1	18.1	6.8
<20	0.02–0.10	82.2	13.6	4.2
<20	>0.10	67.1	27.8	5.1
20–24	<0.02	50	34.7	15.3
20–24	0.02–0.10	59.5	22.5	18
20–24	>0.10	28.6	21.4	50
>24	<0.02	29	35.5	35.5
>24	0.02–0.10	17.2	36.1	46.7
>24	>0.10	46.1	30.8	23.1

Notes: Conditions for the two environmental factors were modified first separately (first six cases) and then together (final nine cases). Empty cells indicate that the variable was not modified.

temperature, but temperature variability alone was not able to explain the variance in cyanobacteria biomass. Rather, a combination of temperature and nutrient availability provide some explanatory power. Similarly, in 18 lakes in Europe over a 23-year period, cyanobacterial biomass increase was statistically linked to longer and stronger stratification (Blenckner et al. 2007).

Water temperature is consistently one of the most important drivers of cyanobacterial blooms but it is interrelated with other factors such as seasonal changes in water column stability, light availability, and nutrient availability (Anneville et al. 2005, Elliott et al. 2005, Wagner and Adrian 2009). In our study, the sensitivity of cyanobacterial abundance to nutrient availability was lower than might be expected, especially compared with outcomes from other modelling and statistical studies (Elliott et al. 2005, Chorus and Schauser 2011). In particular, it was surprising that abundances in the "high hazard" category occurred even at low TP concentrations. However, this result was observed also by Carvalho et al. (2013) analyzing 800 European lakes and it occurred at high temperatures only, which is typically associated with stable stratification. Thus, surface accumulation of blooms may occur at abun-

dances well above those expected from the epilimnetic TP concentrations and water temperature may be a proxy for other important aspects leading to major changes in distribution.

The physical and chemical conditions at the time of sampling are not necessarily those that give rise to the instantaneous observed population but rather it is the immediate past history to which the cyanobacteria have responded (Reynolds 2006). It was necessary in our case to select an appropriate lag time that accounted for both the lagged response of changes in phytoplankton growth and biomass to the environment and for the temporal response of community succession. In our study average meteorological and temperature conditions one week before the bloom event were used to characterize a relatively fast response of biomass. The choice of this time lag could be evaluated in more detail but only with a highly temporally resolved data set that included alignment of physicochemical and cyanobacteria monitoring.

Modelling any ecosystem necessarily demands simplification of the key processes. It is well established that phytoplankton populations have three major requirements for growth: nutrients and light, with temperature playing a moderating effect and mixing influencing position of cells in the water column (Ganf and Oliver 1982, Walsby 1994, Bouterfas et al. 2002). While many physical and chemical processes were not modelled explicitly in our Bayesian network a reasonable prediction of cyanobacterial occurrence was achieved using the readily measured variables of total phosphorus concentration and water temperature. The process of recruitment or germination from akinetes or resting stages is poorly defined and is generally not well represented in most phytoplankton models, with the exception of Hense and Beckmann (2006) and Hense and Burchard (2010). A further compounding factor for predicting cyanobacteria is the spatial variability that occurs with site-specific growth or wind-driven accumulations of cyanobacteria in the leeward part of lakes (Oliver et al. 2012). Despite apparent limitations in representing temporal dynamics, life cycle components, and spatial variability, the Bayesian network was able to accurately predict the probability of cyanobacteria occurring for the different hazard classes in at least 60% of cases.

The performance of a Bayesian network is highly dependent on the data set adopted for the network

TABLE 4. Probabilities of bloom development of CyanoHazard classes for (1) initial conditions (state b for WT and TP); (2) simulated warming by 0.8°C (state c for WT); (3) simulated increase in total phosphorus by 0.01 mg/L (state c for TP).

	Probability of CyanoHazard class bloom (%)				
CyanoHazard class	1) Initial conditions	2) Simulated warming by 0.8 °C	Variation between 1 and 2	3) Simulated increase in TP by 0.01 mg/L	Variation between 1 and 3
High	18.8	41.4	22.6	23.4	4.6
Moderate	42.1	41.4	−0.7	29.9	−12.2
Low	39.1	17.2	−21.9	23.4	−15.7

Note: States b and c refer to Bayesian network in Fig. 3c.

development. An optimal data set for this study would have been a collection of observations including physical variables at daily intervals, and chemical and biological variables at weekly or fortnightly intervals. The optimal resolution to account for chemical and biological variability is difficult to infer and only recently are data from automatic sensors starting to offer some insights (Kara et al. 2012). Ideally, the observations for each lake would also have included several years of observations and lakes would have been equally distributed in space. We organized data to have the maximum number of complete observations to populate the Bayesian network, although the number of complete cases available decreased rapidly when the network complexity (including more variables) was increased. Therefore, the most complex network with nine nodes, potentially has limited predictive ability because it is constrained by the number of suitable observations (271 vs. >1600 used in the three-node network). As observed by Hamilton et al. (2009), predictions become more challenging when few data are available and many variables are included in the network. In our study, it was necessary to balance additions of more explanatory power through adding new variables with the amount of data available.

Bayesian networks are not designed to simulate the evolution in time of particular processes (Pollino and Henderson 2010). To analyze the dynamics of processes, for example how environmental conditions evolve and affect the succession and timing of phytoplankton development, deterministic models are more suitable. By contrast probabilistic models, such as Bayesian networks, are able to associate a particular combination of conditions with a specific event, to estimate the probability of this event occurring. One of their major advantages is that they account for uncertainty. This minimizes the risk of applying management strategies based on incorrect predictions. Accounting for uncertainty in deterministic models is possible, but multiple simulations are needed with a range of different model parameters, often requiring considerable experience of the modeler. To properly express deterministic ecological model predictions, evaluation of physical and biological sources of uncertainty is required (Rigosi and Rueda 2012). The adoption of Bayesian networks may provide an additional tool to answer ecological questions, to evaluate the probability of changes in water quality, to test future scenarios and to establish relevant management procedures. Use of Bayesian networks to analyze and interpret hypotheses and to support decision making has been highlighted previously (Ellison 1996, Castelletti and Soncini-Sessa 2007) and here it has been demonstrated that they can be used to assist with understanding the probability of cyanobacterial hazardous events and potentially supporting decisions relevant to water quality and health risk management.

We were able to adapt a Bayesian network model to account for the effect of climate change when estimating

cyanobacterial risk while also taking into account the interactions between changes in nutrient availability (e.g., representing a modification of land use in a catchment basin) and temperature. A strong dependence on temperature was shown; an increase of 0.8°C for temperatures between 20° and 24°C generated a 5% increase in the probability of hazardous bloom development and a 20% increase in bloom probability occurred after 100 years considering a trend of warming of surface water temperature by 0.037°C/yr (Schneider and Hook 2010). This effect, however, was shown to be strongly regulated by nutrient availability as previously suggested by Brookes and Carey (2011) and recently supported by Rigosi et al. (2014). The Bayesian model outputs not only suggest that regulating total phosphorus availability in the system will help counteract the outcomes of a warming climate but give a quantitative outcome to this hypothesis.

In summary, the Bayesian network was a useful instrument to: explore the interactions between nutrients and temperature simultaneously; estimate the probability of cyanobacterial blooms under warmer conditions and quantify the degree of nutrient reduction that would be required to counteract the effect of an increase in lake water temperature. The simulations provided estimates of how much the total phosphorus concentration should be reduced in order to produce a change in the probability of bloom development equivalent for specific increases in water temperature.

Literature Cited

Anderson, D. M., P. M. Gilbert, and J. M. Burkholder. 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25:704–726.

Anneville, O., S. Gammeter, and D. Straile. 2005. Phosphorus decrease and climate variability: mediators of synchrony in

phytoplankton changes among European peri-alpine lakes. Freshwater Biology 50:1731–1746.

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

Arhonditsis, G. B., C. A. Stow, L. J. Steinberg, M. A. Kenney, R. C. Lathrop, S. J. McBride, and K. H. Reckhow. 2006. Exploring ecological patterns with structural equation modelling and Bayesian analysis. Ecological Modelling 192:385–409.

Blenckner, T., et al. 2007. Large-scale climatic signatures in lakes across Europe: A meta-analysis. Global Change Biology 13:1314–1326.

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

Bouterfas, R., M. Belkoura, and A. Dauta. 2002. Light and temperature effects on the growth rate of three freshwater algae isolated from a eutrophic lake. Hydrobiologia 489:207–217.

Briand, J., C. Leboulanger, J. Humbert, C. Bernard, and P. Dufour. 2004. *Cylindrospermopsis Raciborskii* (Cyanobacteria) invasion at mid-latitudes: selection, wide physiological tolerance or global warming? Journal of Phycology 40:231–238.

Bromley, J., N. A. Jackson, O. J. Clymer, A. M. Giacomello, and F. V. Jensen. 2005. The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. Environmental Modelling and Software 20:231–242.

Brookes, J. D., and C. C. Carey. 2011. Resilience to blooms. Science 334:46–47.

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

Carpenter, S. R., E. H. Stanley, and M. J. Vander Zanden. 2011. State of the world's freshwater ecosystems: physical, chemical and biological changes. Annual Review of Environment and Resources 36:75–99.

Carvalho, L., and A. Kirika. 2003. Changes in shallow lake functioning: response to climate change and nutrient reduction. Hydrobiologia 506–509:789–796.

Carvalho, L., et al. 2013. Sustaining recreational quality of European lakes: minimizing the health risks from algal blooms through phosphorus control. Journal of Applied Ecology 50:315–323.

Carvalho, L., C. A. Miller (nee Ferguson), E. M. Scott, G. A. Codd, P. S. Davies, and A. N. Tyler. 2011. Cyanobacterial blooms: statistical models describing risk factors for national-scale lake assessment and lake management. Science of the Total Environment 409:5353–5358.

Castelletti, A., and R. Soncini-Sessa. 2007. Bayesian networks and participatory modelling in water resource management. Environmental Modelling and Software 22:1075–1088.

Chorus, I., and J. Bartram. 1999. Toxic cyanobacteria in water, a guide to their public health consequences, monitoring and management. World Health Organization, London, UK.

Chorus, I., and I. Schauser. 2011. Oligotrophication of Lake Tegel and Schlachtensee, Berlin. Pages I–IV in I. Chorus and I. Schauser, editors. Analysis of system components, causalities and response thresholds compared to responses of other waterbodies. Federal Environment Agency (Umweltbundesamt), Germany.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117–143.

Clarke, K. R., and R. N. Gorley. 2006. PRIMER v6: user manual/tutorial. PRIMER-E, Plymouth, UK.

Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot, and G. E. Likens. 2009. Controlling eutrophication: nitrogen and phosphorus. Science 323:1014–1015.

Dodds, W. K., W. W. Bouska, J. L. Eitzmann, T. J. Pilger, K. L. Pitts, A. J. Riley, J. T. Schloesser, and D. J. Thornbrugh. 2009. Eutrophication of U.S. freshwaters: analysis of potential economic damages. Environmental Science and Technology 43:12–19.

Dolman, A. M., J. Rucker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: the influence of nitrogen versus phosphorus. PLoS ONE 7:e38757.

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting cyanobacteria dominance in lakes. Canadian Journal of Fisheries and Aquatic Sciences 58:1905–1908.

Elliott, A., and L. May. 2008. The sensitivity of phytoplankton in Loch Leven (U.K.) to changes in nutrient load and water temperature. Freshwater Biology 53:32–41.

Elliott, A., S. J. Thackeray, C. Huntingford, and R. G. Jones. 2005. Combining a regional climate model with a phytoplankton community model to predict future changes in phytoplankton in lakes. Freshwater Biology 50:1404–1411.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6:1036–1046.

Fristachi, A., et al. 2009. Occurrence of cyanobacterial harmful algal blooms workgroup report. Pages 46–103 in H. K. Hudnell, editor. Cyanobacterial harmful algal blooms: state of the science and research needs. Springer Science, New York, New York, USA.

Fujimoto, N., R. Sudo, N. Sugiura, and Y. Inamori. 1997. Nutrient-limited growth of *Microcystis areuginosa* and *Phormidium tenue* and competition under various N:P supply ratios and temperatures. Limnology and Oceanography 42:250–256.

Gallina, N., O. Anneville, and M. Beniston. 2011. Impacts of extreme air temperatures on cyanobacteria in five deep peri-Alpine lakes. Journal of Limnology 70:186–196.

Ganf, G. G., and R. L. Oliver. 1982. Vertical separation of light and available nutrients as a factor causing replacement of green algae by blue-green algae in the plankton of a stratified lake. Journal of Ecology 70:829–844.

Grobbelaar, J. U., and P. S. Stegmann. 1976. Biological assessment of the euphotic zone in a turbid man-made lake. Hydrobiologia 48:263–266.

Haenni, R., J. Romeijn, G. Wheeler, and J. Williamson. 2011. Probabilistic logics and probabilistic networks. Springer Science+Business Media, New York, New York, USA.

Hallegraeff, G. M. 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32:79–99.

Hamilton, G., C. Alston, T. Chiffings, E. Abal, B. T. Hart, and K. Mengersen. 2005. Integrating science though Bayesian belief networks: case study of *Lyngbya* in Moreton Bay. Pages 392–398 in Proceedings of International Congress on Modelling and Simulation 2005, 12–15 December, Melbourne. QUT, Melbourne, Victoria.

Hamilton, G., F. Fielding, A. W. Chiffings, B. T. Hart, R. W. Johnstone, and K. Mengersen. 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: An International Journal 13:1271–1287.

Hamilton, G., R. McVinish, and K. Mengersen. 2009. Bayesian model averaging for harmful algal bloom prediction. Ecological Applications 19:1805–1814.

Hense, I., and A. Beckmann. 2006. Towards a model of cyanobacteria life cycle—effects of growing and resting stages on bloom formation of N_2-fixing species. Ecological Modelling 195:205–218.

Hense, I., and H. Burchard. 2010. Modelling cyanobacteria in shallow coastal seas. Ecological Modelling 221:238–244.

Huber, V., C. Wagner, D. Gerten, and R. Adrian. 2012. To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. Oecologia 169:245–256.

Hudson, J. J., W. D. Taylor, and D. W. Schindler. 2000. Phosphate concentrations in lakes. Nature 406:54–56.

Jensen, F. V. 1996. An introduction to Bayesian networks. Springer-Verlag, New York, New York, USA.

Jöhnk, K., J. Huisman, J. Sharples, B. Sommeijer, P. M. Visser, and A. M. Stroom. 2008. Summer heatwaves promote blooms of harmful cyanobacteria. Global Change Biology 14:495–512.

Kara, E. L., et al. 2012. Time-scale dependence in numerical simulations: assessment of physical, chemical, and biological predictions in a stratified lake at temporal scales of hours to months. Environmental Modelling and Software 35:104–121.

Kosten, S., et al. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Global Change Biology 18:118–126.

Lee, J. H. W., Y. Huang, M. Dickman, and A. W. Jayawardena. 2003. Neural network modelling coastal algal blooms. Ecological Modelling 159:179–201.

Levine, S. N., and D. W. Schindler. 1999. Influence of nitrogen to phosphorus supply ratios and physicochemical conditions on cyanobacteria and phytoplankton species composition in the Experimental Lakes Area, Canada. Canadian Journal of Fisheries and Aquatic Sciences 56:451–466.

Maier, H. R., G. C. Dandy, and M. D. Burch. 1998. Use of artificial neural networks for modelling cyanobacteria Anabaena spp. in the River Murray, South Australia. Ecological Modelling 105:257–272.

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

Martin, J. L., and S. C. McCutcheon. 1999. Hydrodynamics and transport for water quality modelling. Lewis Publishers, Washington, D.C., USA.

McCann, R. K., B. G. Marcot, and R. Ellis. 2006. Bayesian belief networks: applications in ecology and natural resource management. Canadian Journal of Forest Research 36:3053–3062.

Muttil, N., and K. W. Chau. 2006. Neural network and genetic programming for modelling coastal algal blooms. International Journal of Environment and Pollution 28:223–238.

Nalewajko, C., and T. P. Murphy. 2001. Effects of temperature, and availability of nitrogen and phosphorous on the abundance of Anabaena and Microcystis in Lake Biwa, Japan: an experimental approach. Limnology 2:45–48.

Oliver, R. L., D. P. Hamilton, J. D. Brookes, and G. G. Ganf. 2012. Physiology, blooms and prediction of planktonic cyanobacteria. Pages 155–194 in B. A. Whitton, editor. Ecology of cyanobacteria II: their diversity in space and time. Springer Science, New York, New York, USA.

Paerl, H. W., and J. Huisman. 2008. Blooms like it hot. Science 320:57–58.

Pollino, C. A., and C. Henderson. 2010. Bayesian networks: a guide for their application in natural resource management and policy. Landscape Logic, Australian Government Department of the Environment, Water, Heritage and the Arts, Canberra, Australia.

Quinn, J. M., R. M. Monaghan, V. J. Bidwell, and S. R. Harris. 2013. A Bayesian belief network approach to evaluating complex effects of irrigation-driven agricultural intensification scenarios on future aquatic environmental and economic values in a New Zealand catchment. Marine and Freshwater Research 64:460–474.

Read, J. S., D. P. Hamilton, I. D. Jones, K. Muraoka, L. A. Winslow, R. Kroiss, C. H. Wu, and E. Gaiser. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environmental Modelling and Software 26:1325–1336.

Reckhow, K. H. 1999. Water quality prediction and probability network models. Canadian Journal of Fisheries and Aquatic Sciences 56:1150–1158.

Reynolds, C. S. 1997. Excellence in ecology, vegetation processes in the pelagic: a model for the ecosystem theory. Ecology Institute, Oldendorf/Luhe, Germany.

Reynolds, C. S. 2006. The ecology of phytoplankton. Cambridge University Press, Cambridge, UK.

Reynolds, C. S., J. Padisak, and U. Sommer. 1993. Intermediate disturbance in the ecology of phytoplankton and the maintenance of species diversity: a synthesis. Hydrobiologia 249:183–188.

Rigosi, A., C. C. Carey, B. W. Ibelings, and J. D. Brookes. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. Limnology and Oceanography 59:99–114.

Rigosi, A., W. Fleenor, and F. Rueda. 2010. State-of-the-art and recent progress in phytoplankton succession modelling. Environmental Reviews 18:423–440.

Rigosi, A., and F. J. Rueda. 2012. Propagation of uncertainty in ecological models of reservoirs: from physical to population dynamic predictions. Ecological Modelling 247:199–209.

Ryan, E. F., D. P. Hamilton, and G. E. Barnes. 2003. Recent occurrence of Cylindrospermopsis raciborskii in Waikato lakes of New Zealand. New Zealand Journal of Marine and Freshwater Research 37:829–836.

Schneider, P., and S. J. Hook. 2010. Space observations of inland water bodies show rapid surface warming since 1985. Geophysical Research Letters 37:L22405.

Sinha, R., L. A. Pearson, T. W. Davis, M. A. Burford, P. T. Orr, and B. A. Neilan. 2012. Increased incidence of Cylindrospermopsis raciborskii in temperate zones—is climate change responsible? Water Resources 46:1408–1409.

Spiegelhalter, D. J., A. P. Dawid, S. L. Lauritzen, and R. G. Cowell. 1993. Bayesian analysis in expert systems. Statistical Science 8:219–283.

Wagner, C., and R. Adrian. 2009. Cyanobacteria dominance: quantifying the effects of climate change. Limnology and Oceanography 54:2460–2468.

Walsby, A. E. 1994. Gas vesicles. Microbiological Reviews 58:94–144.

Winter, J., A. M. DeSellas, R. Fletcher, L. Heintsch, A. Morley, L. Nakamoto, and K. Utsumi. 2011. Algal blooms in Ontario, Canada: increases in reports since 1994. Lake and Reservoir Management 27:107–114.