Seasonal forecasting of lake water quality and algal bloom risk using a continuous Gaussian Bayesian network

Leah A. Jackson-Blake[1], François Clayer[1], Sigrid Haande[1], James E. Sample[1], S. Jannicke Moe[1]

[1]Norwegian Institute for Water Research (NIVA), 0599 Oslo, Norway

Correspondence to: Leah A. Jackson-Blake (leah.jackson-blake@niva.no)

Abstract

Freshwater management is challenging, and advance warning that poor water quality was likely, a season ahead, could allow for preventative measures to be put in place. To this end, we developed a Bayesian network (BN) for seasonal lake water quality prediction. BNs have become popular in recent years, but the vast majority are discrete. Here, we developed a Gaussian Bayesian network (GBN), a simple class of continuous BN. The aim was to forecast, in spring, mean total phosphorus (TP), chlorophyll-a (chl-a) and water colour and maximum cyanobacteria biovolume for the upcoming growing season (May-October) in lake Vansjø in southeast Norway. To develop the model, we first identified controls on inter-annual variability in seasonally-aggregated water quality. These variables were then included in a GBN and conditional probability densities were fit using observations (≤ 39 years). GBN predictions had R^2 values of 0.38 (cyanobacteria) to 0.75 (colour) and classification errors of 32% (TP) to 17% (cyanobacteria). For all but lake colour, including weather variables did not improve predictive performance (assessed through cross validation). Overall, we found the GBN approach to be well-suited to seasonal water quality forecasting. It was straightforward to produce probabilistic predictions, including the probability of exceeding management-relevant thresholds. The GBN could be sensibly parameterised using only the observed data, despite the small dataset. Developing a comparable discrete BN was much more subjective and time-consuming. Although low interannual variability and high temporal autocorrelation in the study lake meant the GBN performed similarly to a seasonal naïve forecast (where the forecasted value is simply the value observed the previous growing season), we believe the forecasting approach presented could be particularly useful in areas with higher sensitivity to catchment nutrient delivery and seasonal climate and for forecasting at shorter time scales (e.g. daily to monthly). Despite the parametric constraints of GBNs, their simplicity, together with the relative accessibility of BN software with GBN handling, means they are a good first choice for BN development with continuous variables.

1. Introduction

Despite their importance, freshwaters are under intense pressure from human activities. Severe declines in the quantity and quality of habitats and species abundance are widespread, and freshwaters are now one of the most threatened ecosystem types in large parts of the world (Dudgeon et al., 2006; Gozlan et al., 2019; Reid et al., 2019). To try to safeguard freshwater condition, the EU Water Framework Directive (WFD) requires all waterbodies to achieve at least “Good” ecological status by 2027, assessed using a set of indicators of ecosystem integrity (EC, 2003). However, meeting environmental targets is challenging, and despite widespread implementation of measures to improve water quality, 60% of European surface waters were still below “Good” ecological status in 2018 (Kristensen et al., 2018). Harmful cyanobacterial blooms are a particular concern worldwide as they can produce harmful toxins, damage ecosystems, jeopardise drinking water supplies, fisheries and recreational areas, and are becoming more widespread, frequent and intense due to eutrophication and climate change (Huisman et al., 2018; Ibelings et al., 2016; Taranu et al., 2015).

Advance warning, a season in advance, that poor water quality was likely could allow for measures to be put in place to reduce the impacts. For example, water levels could be raised or lowered in flow-regulated waterbodies or more stringent farm management or effluent discharge advice could be issued. Although many cyanobacteria forecasting systems have been developed, the majority predict conditions at most a month in advance or focus on multi-decadal predictions (reviewed in Rousso et al., 2020). Seasonal forecasts, issued with lead times of 1-6 months, could allow for more comprehensive preventative or mitigative measures. Seasonal forecasting is a growing area of research, taking advantage of developments in seasonal climate forecasting, and there are many potential management applications (Bruno Soares & Dessai, 2016). However, seasonal forecasting within the water sector has so far been largely focused on streamflow forecasting, with very limited applications to lake water quality forecasting. The focus of the WATExR project, a European Union project funded by the European Research Area for Climate Services (ERA4CS), was to help address this gap by developing pilot seasonal forecasting tools for lake water quality and ecology. Tools were co-developed with water managers at five catchment–lake case study sites, with four in Europe and one in South Australia (Jackson-Blake et al., 2022). Tools linked seasonal climate forecasts with models for predicting river discharge, lake water level and water temperature (Mercado-Bettín et al., 2021), water quality, algal bloom risk, and fish migration. Here, we describe the model developed to forecast lake water quality at one of the case study sites, Lake Vansjø in Norway.

A multitude of potential methods exist for water quality modelling and forecasting. Here, we adopt a Bayesian network (BN) approach. BNs are a type of probabilistic multivariate model which is well suited to environmental modelling, risk assessment and forecasting (Aguilera et al., 2011; Kaikkonen et al., 2021; Uusitalo, 2007). In brief, BNs are graphical models in which the joint probability distribution among a set of variables $X = [X_1, \ldots X_n]$ is represented in terms of: (1) a directed acyclic graph, where each vertex (or node) represents a variable in the model and an edge (or arc) linking two variables indicates statistical dependence, and (2) conditional distributions for each variable X_i, $p(X_i|p_a(X_i))$, given the probability distribution $p_a(X_i)$ of any parent nodes, which quantify the strength and shape of dependencies between variables (Pearl, 1986). In recent years BNs have become popular in a broad range of environmental modelling disciplines, including modelling lake water quality and algal bloom risk (e.g. Couture et al., 2018; Gudimov et al., 2012; Rigosi et al., 2015; Shan et al., 2019; Williams & Cole, 2013). Particular strengths in terms of our seasonal forecasting aims are that, as nodes are modelled using probability distributions, risk and uncertainty can be estimated easily and accurately compared to many other modelling approaches. They can thus be powerful tools to assess the probability of events (e.g. WFD ecological status class). They are also well suited for communicating and visualizing the model to end users and it is easy to update the model given new data. Other benefits include the ability to model complex systems in a quick and efficient way, to combine data and expert knowledge, easy handling of missing values, and the potential to be used for inference as well as prediction.

BNs were originally designed to deal with discrete data. Relationships between nodes in discrete BNs can be non-linear and complex, thereby allowing for the full power of BN analysis, and the vast majority of environmental BN models are discrete (Aguilera et al., 2011). Any continuous variables must first be discretized, but this involves an information loss as discretization can only capture the rough characteristics of the original distribution. In addition, discretization choices (number of intervals and division points) affect BN results (e.g. Nojavan et al., 2017) and their interpretation (Qian & Miltner, 2015). In practice, it is usually necessary to restrict the number of intervals, often to just two or three classes, as the more intervals, the more data are needed to parameterise the model meaningfully (Hanea et al., 2015). Such restrictions mean it then becomes difficult to capture complex relationships, thereby diminishing the theoretical benefits of using a discrete network (Uusitalo, 2007). Continuous BNs, by contrast, represent continuous variables using continuous statistical distributions or equations, and therefore avoid the need for discretization. Hybrid BNs, which include both discrete and continuous nodes, have similar benefits.

In recent years, much focus on continuous networks has been aimed at developing algorithms for non-parametric networks, i.e. continuous networks which are not limited by assumptions about the nature of the statistical distribution of continuous variables (Marcot & Penman, 2019). However, Gaussian BNs (GBN) are a long-established, simple and powerful class of continuous BN, and are often the only type of continuous node available in commonly-used BN software (e.g. Bayes server, bnlearn, Hugin). In GBNs, each random variable is defined by a Gaussian distribution and variables are linearly related to their parents (Geiger & Heckerman, 1994; Shachter & Kenley, 1989). In some situations these parametric constraints may be overly limiting but, when this approximation is appropriate, GBNs may be preferable over discretization. Despite the potential benefits, the use of continuous BNs in environmental modelling is rare. In a review of papers published over the period 1990-2010, Aguilera et al. (2011) found only 6% included continuous or hybrid data, and we could only find 9 more recent examples in the literature (web of science search in November 2021 with terms [environmental AND modelling* AND "Bayesian network" AND continuous], with manual sorting of results).

Overall aims of the paper were therefore: (1) to develop a model for seasonal forecasting of lake water quality, and (2) to demonstrate the use of a continuous GBN, instead of more traditional discrete BN approaches. Our case study site is the western basin of lake Vansjø, a shallow mesotrophic/eutrophic lake in southeast Norway. A number of BN models have previously been applied in the lake (Barton et al., 2008; Couture et al., 2018; Couture et al., 2014; Moe et al., 2019; Moe et al., 2016), but these were all discrete meta-models, i.e. the underlying network nodes were 'response surfaces' summarising a combination of process-based model simulations, statistical relationships, expert opinion and/or data distributions, and the studies were focused on the longer-term impacts of climate, land use and land management change. Here, the aim was to provide medium-term forecasts to support lake management, by developing a model able to predict, in spring of a given year, water quality for the coming growing season (May – October), including the probability of lying within WFD ecological status classes for TP, chl-a and cyanobacteria. We also forecast lake colour, as elevated lake organic matter content (and associated colour) can cause a number of problems for drinking water treatment (see e.g. Matilainen et al., 2010). To develop the model we took a data-driven approach: we first used exploratory statistical analyses to identify the main controls on interannual variability in lake water quality, then combined the results of this with domain knowledge to develop the GBN, and finally parameterised it using 39 years of data. We then explored the sources of predictability and the importance of weather variables by comparing predictive performance of GBNs with different model structures within a cross validation scheme. We also compare GBN predictive ability to a comparable discrete BN and to a simple benchmark model.

2. Methods and data

2.1. Case study site

Lake Vansjø is a large lake in southeast Norway (59°24′N 10°42′E, 25 m asl), with a highly agricultural catchment by Norwegian standards (15% of the 690 km^2 catchment is agriculture) with clay- and P-rich soils. The lake has two main basins, Storefjorden in the east (24 km^2) and Vanemfjorden in the west (12 km^2) (Fig. 1). The largest input is the Hobøl River (catchment area 301 km^2), which enters Storefjorden, and then water flows from Storefjorden to Vanemfjorden through a narrow channel and from Vanemfjorden through Moss River towards the Oslo Fjord (Fig. 1). Over the period 1989-2018, catchment mean annual air temperature was 7.2 °C and annual precipitation was 992 mm yr^{-1}.

Here, we focus on Vanemfjorden, which is shallower (mean depth 3.8 m, max depth 19 m) and more susceptible to eutrophication and cyanobacterial blooms than Storefjorden, due to stronger interactions between the water column and the P-rich lake sediments and a more agricultural local catchment. Vanemfjorden has a relatively short residence time (0.21 years) and the water column remains oxygenated throughout the year. Vanemfjorden has a long history of eutrophication and is usually in WFD 'Moderate' ecological status for mean growing season TP (> 20 µg/l), chl-a (> 10.5 mg/l) and maximum cyanobacteria (> 1.0 mg/l) (Skarbøvik et al., 2021). Vanemfjorden suffers from toxin-producing cyanobacterial blooms and bathing bans were in place during much of the early 2000s (Haande et al., 2011).

The outlet of Vanemfjorden is dammed and lake water level is regulated for hydropower, recreation, and flood protection. There is a management opportunity for the dam operators to adjust the water level in advance of an anticipated wet, dry or hot season if problematic water quality were expected, whilst the local catchment management group (Morsa), responsible for WFD implementation, are interested in seasonal water quality forecasts to inform their management plan, in particular preparedness for cyanobacterial blooms.

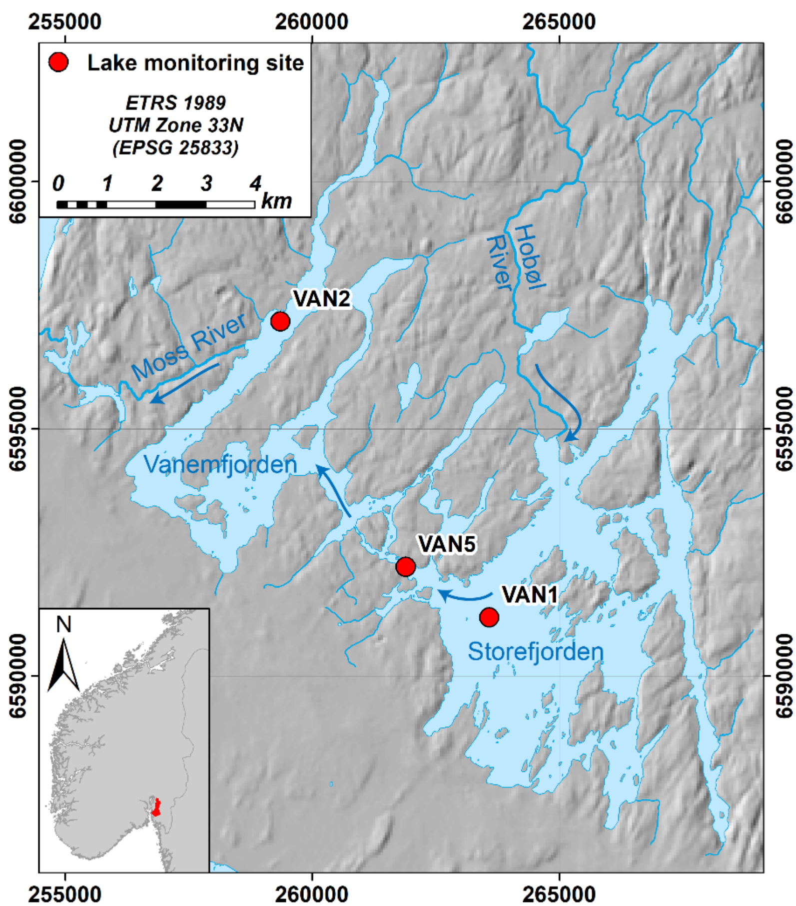


Figure 1. Lake Vansjø in southeast Norway, showing the two main basins: the larger eastern basin (Storefjorden) and Vanemfjorden (the study basin), which has its outlet at Moss River. The two basins are connected by a narrow channel. The largest tributary to Lake Vansjø is the Hobøl River. Main NIVA monitoring sites are shown and arrows show the dominant flow directions. Here, we use data from VAN2.

2.2. Overview of the workflow

The aim was to develop a seasonal forecasting model capable of producing probabilistic forecasts, issued in spring of a given year, of expected growing season (May-October) mean concentrations of TP and chl-a and maximum cyanobacteria biovolumes, as used in WFD status classification for Norwegian lakes (Vanndirektivet, 2018). Mean lake colour was also forecast, both because it is of interest for drinking water treatment, and because it may influence algal biomass by affecting nutrient and light conditions (Bergström & Karlsson, 2019; Carpenter et al., 1998).

The model development and assessment workflow consisted of the following steps:

1. *Feature generation*: Data pre-processing and temporal aggregation to derive an array of potential explanatory variables (or features, in machine learning parlance).
2. *Feature selection*: Exploratory statistical analyses to identify key features, using a combination of feature importance analysis using regression trees, correlation coefficients and scatterplots. Process knowledge was used as the final selection criteria.
3. *BN development*: selected explanatory variables were incorporated into a GBN, using process knowledge to define the structure. Data from the study site were then used to fit the GBN parameters.
4. *Discrete BN development*: A discrete BN was also developed for comparison, using discretized data and the same structure as the GBN.
5. *BN cross-validation and evaluation*: Selection of the most appropriate GBN structure for each target variable, with a particular focus on any added value from including weather variables, and comparison to the discrete BN.
6. *Benchmarking*: Comparison of GBN predictive skill to a simple benchmark model, a seasonal naïve forecaster.

All pre- and post-processing was carried out in the Python programming language. BN development and cross-validation were carried out using the bnlearn R package (Scutari, 2009; Scutari & Ness, 2012). Scripts and data are available (see Section ‘Code and data availability’).

2.3. Data and temporal aggregation

Meteorological, river flow, river chemistry and lake chemistry data were used to derive potential explanatory variables. Precipitation and air temperature were derived from the seNorge 1 km^2 gridded data (Lussana et al., 2019), averaged over the whole catchment. Wind speed data were from the met.no monitoring location at Rygge airport, by the southern edge of the lake. Hobøl River discharge is measured hourly by NVE at Høgfoss and was aggregated to a daily sum. TP concentration data from the Hobøl River at Kure were downloaded from Vannmiljø (https://vannmiljo.miljodirektoratet.no/, last accessed 01/11/2021).

Lake water quality data were from the surface 0-3 m from monitoring point Van2 (see Fig. 1). TP, chl-a and colour data were downloaded from Vannmiljø whilst cyanobacteria biovolume was provided by NIVA (pers. comm). NIVA colour data were patchy over the period 1998-2007. However, colour is also monitored by Movar, the local drinking water company, and data were obtained for the period 2000-2012 (pers. comm.). Despite different sampling locations and depths (Movar monitoring is in Storefjorden at 20 m depth), the two datasets were highly correlated and from the same distribution. We therefore patched the series together, making maximum use of the higher-frequency MOVAR data: NIVA data were used pre-1999, Movar data from 1999-2012 and NIVA data from 2013. Cyanobacteria monitoring began in 1996, whilst all other variables were monitored from 1980. Prior to 2004, sampling took place 6-8 times a year during May/June to September/October. From 2005, the period changed to mid-April to mid-October, and with higher frequency (fortnightly for cyanobacteria, weekly for other variables between 2005 and 2014 and fortnightly thereafter). The number of samples per growing season therefore varies considerably throughout the period 1980-2018, from 5 to 10 per year until 2004, increasing to around 25 (TP, chl-a, colour) until 2013, and then decreasing to around 12. Monthly and seasonally-aggregated values pre-2005 are therefore based on substantially fewer data points.

Lake TP concentration in Vanemfjorden is fairly constant throughout the growing season and is almost always in the range 25-40 ug/l. Meanwhile, Hobøl river TP concentrations are almost always above this, around 40-140 ug/l. Chl-a and cyanobacteria biovolume tend to peak in July or August. Lake colour is highest in spring and winter and decreases through summer and autumn.

The aim was to predict the WFD status class of a number of key water quality parameters, which in Norwegian lakes are assessed using average or maximum values over the whole growing season (May-Oct) (Solheim et al., 2014). Daily data were therefore truncated to the growing season (May-Oct) and were aggregated over this period by calculating seasonal means, sums, counts or maxima. This 6-monthly aggregated data was then used in all subsequent analyses. Time series for the four lake water quality variables of interest and a number of potential explanatory variables, aggregated over the summer growing season, are shown in Fig. 2. Interannual variability in TP is low, aside from a general decline since around 2001. Chl-a is more variable although longer-term trends still dominate, with an increase until around 1995, high values during 1995-2006, and decreasing thereafter. Cyanobacteria was variable until 2008 and has been low since. There is a step change increase in lake colour between 1997 and 1999. Lake colour has been increasing across Scandinavia over recent decades, so this may be real (de Wit et al., 2016), but it may also be due to e.g. a change of labs or methods, but this could not be confirmed due to a lack of metadata. Some broad-scale trends are also apparent in the potential explanatory variables. Growing season mean air temperature is generally between 12 and 14°C, but was somewhat higher after 2005. Mean wind speed was highest earlier in the period in the 1980s, lowest around 2006-2008, and increased thereafter. This increase over the last decade appears to be mostly due to a lack of calm wind days, and is observed at other nearby meteorological stations (e.g. Skarpsborg). Precipitation shows high variability, but was generally lower in the first half of the study period.

Temporal aggregation over the whole growing season, although of WFD-relevance, is coarse and may miss causative relationships. We therefore also carried out finer-scale aggregation, to check and expand on the results obtained from the 6-monthly analyses (see Appendix A).

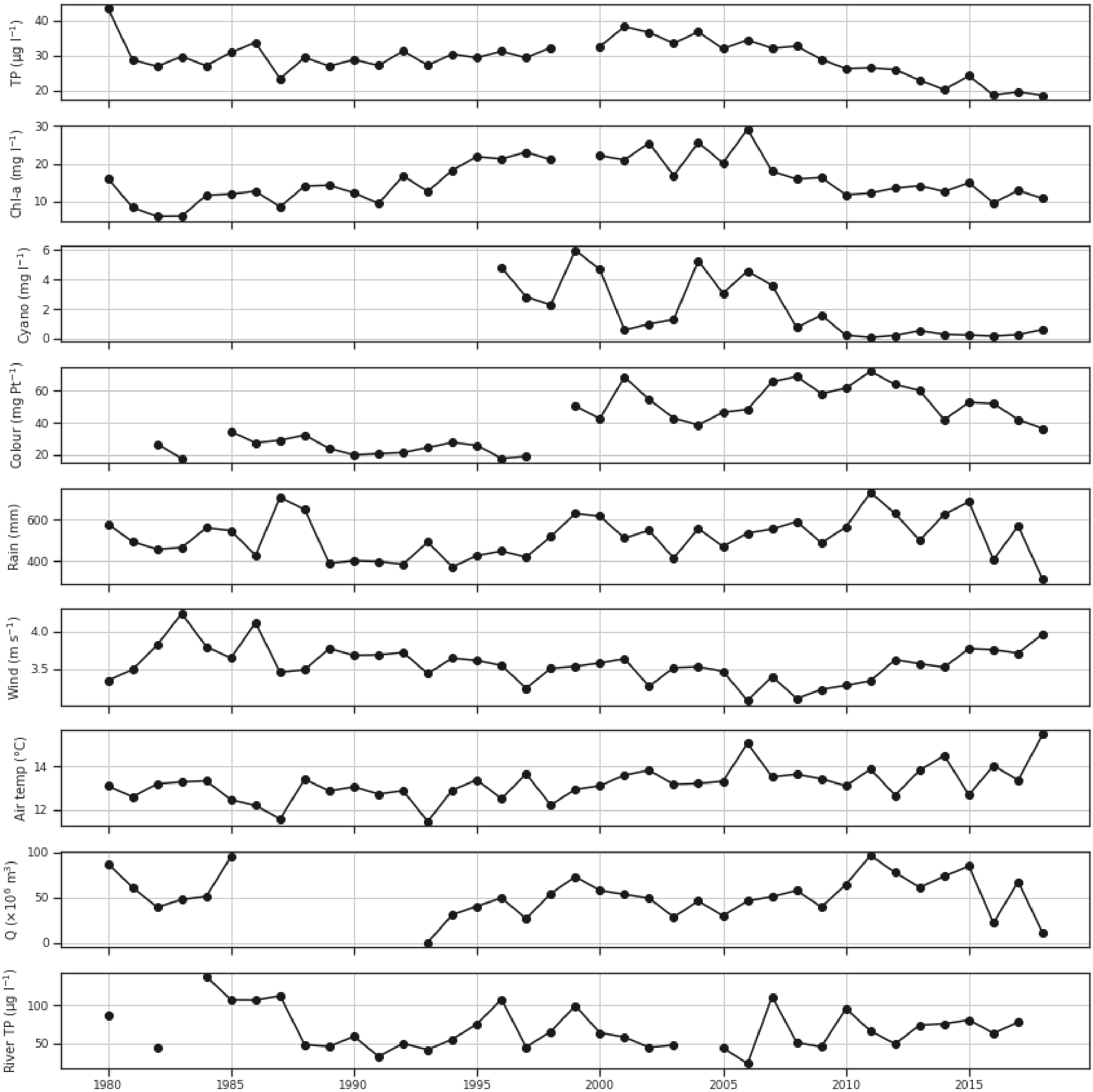


Figure 2. Time series for Lake Vansjø of growing season (May-Oct) mean concentrations of lake chl-a, total phosphorus (TP) and colour, seasonal maxima of cyanobacteria biovolume, seasonal mean wind speed, air temperature and Hobøl River TP concentration, and seasonal sums of rainfall and discharge (Q).

2.4. Feature generation

To issue a forecast for seasonally-aggregated summer lake water quality, we need to first understand the key factors controlling inter-annual variability. Lake TP concentration and colour may be controlled by delivery from the surrounding catchment, interaction with lake sediments, lake stratification and mixing (Søndergaard et al., 2013; Welch & Cooke, 2005). For algal biomass and harmful algal blooms, the right combination of environmental conditions can lead to bloom formation, including sufficiently high nutrient concentrations, in particular P (e.g. Heisler et al., 2008; Lürling et al., 2018; Stumpf et al., 2012), temperature (e.g. Kosten et al., 2012; Paerl & Huisman, 2009; Robarts & Zohary, 1987), light intensity (e.g. Kosten et al., 2012; Merel et al., 2013), and a stable water column (e.g. Huber et al., 2012; Yang et al., 2016). The relative importance of different drivers varies according to lake type, with nutrients often providing a dominant control in polymictic lakes (shallow lakes whose waters frequently or continuously mix vertically throughout the ice-free period), whilst dimictic lakes (which fully mix vertically twice a year) are generally more sensitive to climatic variables through their effect on water column stability (Taranu et al., 2012).

To determine the key explanatory variables in our study site, we generated a set of potential variables (or features) for each of the lake water quality variables of interest. As the aim was to produce a seasonal forecasting model, our choice of variables was somewhat limited to data which would be available or could be readily modelled at the time the forecast was issued. Historic lake water quality observations and weather were therefore included, as were interrelationships between growing season water quality variables, as BNs allow for multiple variables to be predicted at the same time. Growing season weather variables and features relating to the delivery of water and TP from the catchment were also generated. For an operational seasonal forecasting model, these would need to be obtained from external forecasting efforts (e.g. seasonal climate forecasts, or catchment models driven by seasonal climate forecasts). For these variables, we had the choice of using either observed historic data or model-derived hindcasts in our BN model development. We decided to use real observed data, to enable us to assess whether variables were genuinely important using best-available data, but see Section 4.1.2 for a discussion of the use of simulated data instead. Some potentially relevant features (e.g. water temperature and water column stability indices) were not included, as for operational forecasting these would need to be produced by a chain of models (seasonal climate – catchment hydrology – lake) or by adding latent variables to the GBN, both of which were thought to be too complex for the current workflow. In addition, these variables should be proxied by other variables that were included in the feature set.

After choosing the variables to include, features were generated for the current May-October growing season, for the previous winter (the November to April six-month period prior to the current season) and the previous year's growing season. Overall, we generated up to 29 potential explanatory variables, depending on the response variable (Table 1). Features were derived for the period 1981 – 2018. Depending on the number of years with missing data, this gave 39 years of data for TP and chl-a, 36 for lake colour and 24 for cyanobacteria.

Table 1: Potential explanatory variables (features) for each of the four dependent variables. The temporal aggregation period is given relative to the forecast issue date in spring of the current year, *y*.

Dependent variable	Feature name	Description	Temporal aggregation period
Chl-a, cyano	TP	Mean lake TP concentration (µg/l)	Current growing season (May – Oct), year *y*
Chl-a, cyano	Colour	Mean lake colour (mg Pt/l)	
Cyano	Chl-a	Mean lake chl-a concentration (mg/l)	
TP, chl-a, cyano	TP river	Mean TP concentration in the Hobøl River (µg/l)	
All	Rain sum	Precipitation sum (mm)	
	Rain days	Count of rain day (daily precipitation ≥ 1mm)	
	Rain intense	Count of intense rain days (daily precipitation ≥ 10 mm)	
	Q	Inflow discharge sum (10^6 m^3)	
	Temp	Mean of daily mean temperature (°C)	
	Wind speed	Mean of daily mean wind speed (m/s)	
	Wind < P20	Count of days when daily mean wind speed < 20^{th} percentile (2.0 m/s)	
	Wind < P40	Count of days when daily mean wind speed < 40^{th} percentile (2.9 m/s)	
	Wind > P60	Count of days when daily mean wind speed > 60^{th} percentile (3.8 m/s)	
	Wind > P80	Count of days when daily mean wind speed > 80^{th} percentile (4.8 m/s)	
All	Rain sum (W)	Precipitation sum (mm)	Previous winter (Nov year *y*-1 to April year *y*)
	Rain days (W)	Count of rain days (daily precipitation ≥ 1mm)	
	Rain intense (W)	Count of intense rain days (daily precipitation ≥ 10 mm)	
	Q (W)	Inflow discharge sum (10^6 m^3)	
	Temp (W)	Mean of daily mean temperature (°C)	
	Wind speed (W)	Mean of daily mean wind speed (m/s)	
	Wind < P20 (W)	Count of days when daily mean wind speed < 20^{th} percentile (2.0 m/s)	
	Wind < P40 (W)	Count of days when daily mean wind speed < 40^{th} percentile (2.9 m/s)	
	Wind > P60 (W)	Count of days when daily mean wind speed > 60^{th} percentile (3.8 m/s)	
	Wind > P80 (W)	Count of days when daily mean wind speed > 80^{th} percentile (4.8 m/s)	
All	Temp (PS)	Mean air temperature (May-Oct; °C)	Previous year's growing season (May – Oct, year *y*-1)
TP, chl-a, cyano	TP (PS)	Mean lake TP concentration (µg/l)	
Chl-a, cyano	Chl-a (PS)	Mean lake chl-a concentration (mg/l)	
Colour, chl-a, cyano	Colour (PS)	Mean lake colour (mg Pt/l)	
Cyano	Cyano (PS)	Mean lake cyanobacterial biovolume (mg/l)	

2.5. Feature selection

Having generated a list of potential explanatory variables for each target variable, we then carried out exploratory statistical analyses to select the features to include in the GBN, using a combination of:

1. *Ranked correlation coefficients*: As a first screening, we used ranked absolute Pearson's correlation coefficients to highlight potentially important features for each dependent variable.
2. *Feature importance*: We used a machine learning approach to assess feature importance, using random forests implemented using the Scikit-Learn python package (Pedregosa et al., 2011). Random forests use bootstrapping to partition the data used by each tree and data not included in each bootstrap sample are used to perform internal validation. We used the out-of-bag (OOB) score and importance scores to rank feature importance. We used recursive feature elimination to try to find the best random forest regressor model using subsets of the available features. This is similar to stepwise regression, but uses cross-validation to avoid overfitting, rather than traditional significance testing, and in this case we used out-of-sample R^2 to measure performance. Random forests have a number of hyperparameters that can be tuned to improve performance. The most important are the number of trees in the forest (n_estimators) and the size of the random subsets of features to consider when splitting a node (max_features). We selected values for these by plotting the OOB error rate (1 – OOB Score) as a function of n_estimators for various choices of max_features.
3. *Visual evaluation of relationships*: Scatterplot matrices were used for a visual check of whether relationships appeared to be linear and for independence between explanatory variables (required for unconnected nodes in a BN).
4. *Process understanding*: Finally, we excluded explanatory variables where we did not think there were plausible physical mechanisms underpinning the relationship.

2.6. Bayesian network development and use in prediction

We first defined the BN structure manually, using results of the exploratory feature selection and process-knowledge, to ensure realistic causative relationships between nodes. This structure was then used in both the continuous Gaussian BN and a discrete BN.

2.6.1. Gaussian Bayesian Network development

As mentioned in the introduction, Gaussian Bayesian networks (GBNs) are a powerful class of continuous BNs in which all nodes are continuous and conditional probability distributions (CPDs) are linear Gaussians, which together define a joint Gaussian. Parent nodes therefore have normal distributions with mean μ and variance σ^2. Gaussian CPDs of child nodes have a mean which is a linear combination of the parent nodes (with intercept β_0 and coefficients β_n). To meet the normality requirement of GBNs, we transformed the cyanobacteria data, which were right skewed with many zeros, by applying a box cox transformation ($y^* = (y^\lambda - 1)/\lambda$ with $\lambda = 0.1$ to give a fairly symmetrical distribution). Predictions for cyanobacteria were then transformed back to the original data scale using bias-adjustment back-transformation (see Chapter 3.2, Hyndman & Athanasopoulos, 2018). Normality tests were carried out for all variables using scipy.stats (based on D'Agostino & Pearson, 1973). High p values (> 0.2) were found for all but lake colour and transformed cyanobacteria ($p = 0.04$ for both). A step change in lake colour is seen around 1998 (Fig. 2) suggesting the distribution of lake colour may be bimodal. The normality assumption was therefore not invalidated at a 1% significance level, but would have been at a 5% level. Coefficients were then derived for the CPDs at each node using maximum likelihood estimation.

BNs can be used for prediction, our primary aim, by calculating a probability distribution over the variable(s) whose value we would like to know, given information (evidence) we have about some other variables. Predictions obtained using GBNs contain a mean and a variance, and here predictions were obtained by averaging likelihood weighting simulations using a subset of nodes as evidence. The predicted value is then the expected value of the conditional distribution. We chose the evidence nodes based on those nodes which could be updated whenever a forecast was produced, using historic data or future forecasts (i.e. observed water quality from the previous summer or forecasted meteorological conditions). As well as predicting absolute values, we also estimated the probability of exceeding a management-relevant threshold for each water quality variable (Table 2).

Table 2: Management-relevant thresholds used for predicting the probability of lake water quality variables lying within a certain water quality class. The classification is summarised as low concentration (L) and high concentration (H) classes, which translate to a WFD-relevant classification as described. WFD is the Water Framework Directive.

Variable	Low/high concentration class boundary	Relationship between concentration class and WFD class	Rationale
TP	29.5 µg/l	Low = Upper Moderate High = Lower Moderate	Almost all observations were within the Moderate WFD status class, so we used the mid-point of this class as the threshold.
Chl-a	20.0 mg/l	Low = Moderate or better High = Poor or worse	WFD Moderate/Poor boundary
Cyano	1.0 mg/l	Low = Good or better High = Moderate or worse	WFD Good/Moderate boundary
Colour	48 mg Pt/l	Not applicable	Upper tercile (66th percentile)

2.6.2. Discrete Bayesian network development

Finally, we developed a discrete BN for comparison to the GBN. To do this, we first discretized the data, opting for just two classes for most variables, given the small sample size. The exception was colour_prevSummer, where we used three classes, given a strong relationship between lake colour in the previous and current growing season (Sect. 3.1). We used management-relevant thresholds to discretize the current growing season lake TP, chl-a, cyanobacteria and colour (Table 2). For all the other variables, including lake observations from the previous summer, as we were not constrained by having to discretize into management-relevant classes, we used regression trees to identify the optimal splitting points, to improve the chances of identifying relationships between nodes in the BN. For each dependent variable (TP, chl-a, cyano, colour), we built a regression tree for each explanatory – dependent variable pair in turn, and then used the first split point in the tree as the boundary for discretizing that explanatory variable. For wind speed, this resulted in highly unbalanced class sizes so we instead used the median. The following boundaries were used: TP (PS): 29.5 mg/l, chl-a (PS): 16.8 mg/l, colour (PS): 32.6 and 61.0 mg Pt/l, rain sum: 497 mm, wind speed: 3.56 m/s. Note that the different discretization methods used for current vs previous year's growing season water quality means that the two variables are classified differently, despite the same underlying data. The resulting classes were relatively well balanced.

We then fitted the CPTs using Bayesian posterior estimation with uniform priors. Including priors helps avoid overfitting, a common problem with maximum likelihood estimation (mle, where CPTs are fitted just using relative frequencies), particularly with small sample sizes when the data may not be representative of the underlying distribution. In our case, priors can be thought of as pseudo state counts added to the actual counts before normalization. The uniform priors were defined by the imaginary sample size (iss), whereby the pseudo counts are the equivalent of having observed iss uniform samples of each variable and each parent configuration. The higher the iss, the stronger the effect of the prior on the posterior parameter estimates, whilst with iss = 0, the method becomes mle. The iss parameter thus specifies the weight of the prior compared to the sample and therefore controls the smoothness of the posterior distribution. A common rule of thumb is to use a small non-zero iss to avoid zero entries. However, we experimented with larger values of iss (from 1 to 50), to avoiding overfitting. We did this using a trial-by-error process. During each iteration, we examined the CPTs for spurious relationships and checked the predictive error of the network through cross validation (see Section 2.7.1). We found that an iss of 15 was the smallest value where the majority of unexpected CPT behaviour was smoothed out, without compromising on predictive performance.

2.7. BN validation and assessment

We then explored the most appropriate GBN model structure and assessed its predictive performance using three methods: (1) cross validation, carried out on several parts of the network separately and including comparison to the discrete BN, (2) goodness of fit of the whole network compared to observations, and (3) comparison to a simple benchmark model.

2.7.1. Cross-validation of sub-networks

The ability to carry out cross-validation (CV) is a benefit of using bnlearn compared to many graphical BN packages, as it is possible to assess the expected performance of the network for out-of-sample prediction, and to compare different models to robustly assess whether certain nodes and arcs are improving predictive power. Here, we used CV to compare the predictive performance of GBNs with and without meteorological nodes and to compare the GBN and the discrete BN. We used leave-one-out cross-validation, which should produce unbiased skill score estimates even with small sample sizes (e.g. Wong, 2015). In short, the cross-validation was repeated for each predicted node (chl-a, cyano, TP, colour) and involved the following steps. One year of data is left out at a time, the BN is fit using the remaining years of data, and is then used to predict the target node for the left out year. The procedure is repeated for all years, producing a single time series of predictions. These are then compared to observations to generate skill scores. As the main aim was prediction, we used posterior predictive correlation (reported as R^2), mean square error (MSE) and classification error (the proportion of the time the classification was incorrect) as GBN skill scores, and just classification error for the discrete BN. The cross-validation is stochastic and was run a default 20 times and the mean of skill scores were calculated.

Cross-validation requires complete data for all variables and years. For most variables there were few gaps, so we filled up to one-year gaps by interpolation or backward/forward filling. However, cyanobacteria was only measured from 1996, whilst other variables were measured from 1980. Rather than dropping all data prior to 1996, which would result in a large loss of training data for TP, chl-a and colour, we instead split the network into a number of smaller networks for the target variables, and cross validated each of these in turn (see Section 3.3.1).

2.7.2. Goodness-of-fit of the whole network

Splitting the BN up into smaller sub-networks is likely to result in a loss of predictive power, so cross-validation could not be used to assess the expected predictive performance of the whole network. Instead, we assessed the performance of the whole network, trained on all data, by simply calculating goodness-of-fit of predictions using the GBN with and without weather nodes. For this, we used correlation, MSE and classification error, as during cross-validation, and bias (mean of (predicted – observed)). We also calculated the Matthew's correlation coefficient (MCC), to provide additional information on how well the WFD status class was predicted. MCC is in the range 0 (no skill) to 1 (perfect skill), and has been shown to be a truthful score for evaluating binary classifiers (Chicco & Jurman, 2020). As the training and evaluation data were the same, this may produce an optimistic assessment of model performance.

2.7.3. Comparison to a benchmark model

Some extremely simple forecasting methods can be highly effective. As a final test, we compared predictive performance of the GBN to a seasonal naïve forecast (Hyndman & Athanasopoulos, 2018). In this case, the seasonal naïve forecast for the current growing season is simply the observed value from the previous year's growing season.

3. Results

3.1. Feature selection

For lake TP concentration, key features identified were TP concentration from the previous growing season and, to a lesser extent, wind-related features (Tables 3 and 4). A regression tree model using just the previous summer's TP (TP (PS)) and the number of calm winter wind days (wind < P20 (W)) had an out-of-bag (OOB) score of 0.35, slightly higher than when all features were included. Temporal autocorrelation in lake TP concentration is highly plausible. It is however less clear whether the negative correlation with wind speed is causative. We might expect windier conditions to increase sediment resuspension and result in higher TP concentrations (Hanlon, 1999), but in fact higher TP was associated with calmer weather (Fig. 3). A positive relationship between the previous summer's TP and winter wind (Fig. 3), together with analyses using monthly aggregated data (Appendix A), suggest the relationship may not be causative. Wind was not therefore selected for TP.

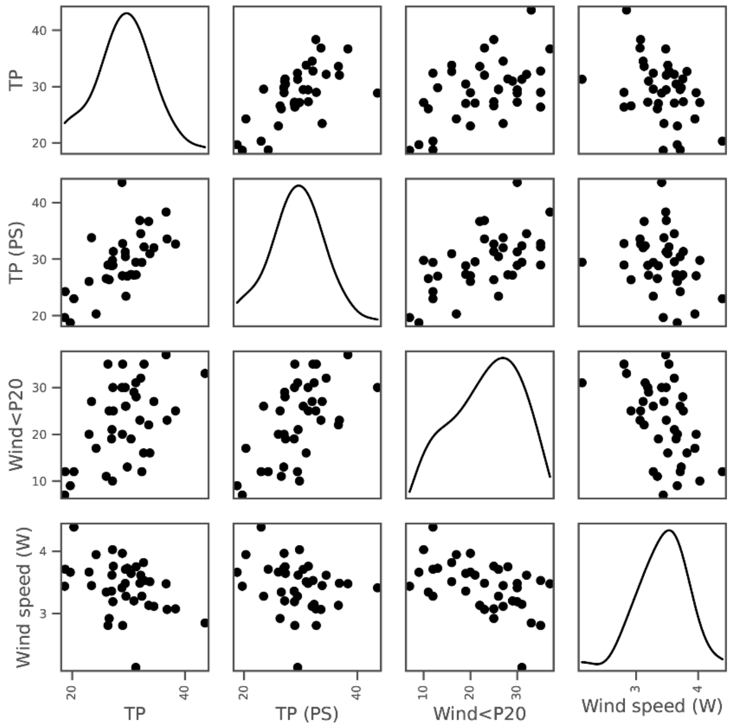


Figure 3: Relationships between seasonal mean lake TP concentration and potential explanatory variables of interest, including lake TP the previous summer (PS), number of days when daily mean wind speed < 20th percentile (wind<P20), and mean winter (Nov-April) wind speed. Density plots estimated using kernel density estimation (kde) are shown along the diagonal.

Table 3: Pearson's correlation coefficients (R) for the four dependent variables (only |R| > 0.40 are shown). See Table 1 for a description of the variables.

TP		Chl-a		Cyanobacteria		Colour	
Variable	R	Variable	R	Variable	R	Variable	R
TP (PS)	0.65	Chl-a (PS)	0.65	Chl-a	0.77	Colour (PS)	0.85
Wind < P20	0.51	TP	0.58	TP	0.58	Rain sum	0.53
Wind < P20 (W)	0.44	Wind < P40	0.41	Chl-a (PS)	0.56	Rain intense	0.46
Wind speed (W)	-0.40	Wind > P80	-0.49	Cyano (PS)	0.55	Q	0.45
		Wind speed	-0.51	TP (PS)	0.49	Temp (PS)	0.43
		Wind > P60	-0.51	Colour	-0.44	Wind > P60	-0.45
				Colour (PS)	-0.50	Wind speed	-0.46
						Wind > P80	-0.47

Table 4: Summary of feature importance analysis for each dependent variable. The out-of-bag (OOB) score gives the overall performance of the random forest regressor model. This was done for a variety of feature subsets, including all available features (All), features included in the best random forest regressor model, identified by recursive feature elimination (Optimum) and the feature subset selected for further BN development (Selected). See Table 1 for a description of the features.

Target variable	Feature subset	Feature	Importance scores	OOB
TP	All	TP (PS)	0.37	0.29
		Wind < P20 (W)	0.15	
		All others	< 0.08	
	Optimum (6)	TP (PS)	0.43	0.41
		Wind < P20 (W)	0.21	
		All others	< 0.12	
	Selected	TP (PS)	1	0.06
Chl-a	All	Chl-a (PS)	0.30	0.48
		TP	0.18	
		Wind speed	0.06	
		All others	< 0.05	
	Optimum and selected	Chl-a (PS)	0.41	0.49
		TP	0.34	
		Wind speed	0.24	
Cyano	All	Chl-a	0.14	0.31
		Colour	0.08	
		All others	< 0.07	
	Optimum	Chl-a	1	0.34
	Selected	Chl-a	0.63	0.55
		Colour	0.37	
Colour	All	Colour (PS)	0.73	0.64
		All others	< 0.05	
	Optimum	Colour (PS)	0.79	0.67
		Wind < P20	0.12	
		Rain sum	0.09	
	Selected	Colour (PS)	0.85	0.57
		Rain sum	0.15	

For chl-a, strongest correlations were with chl-a the previous summer and lake TP concentration (Table 3). Otherwise, the only correlation coefficients above 0.4 were with wind-related features, including a negative relationship with mean wind speed. This was partly supported by the feature importance analysis and a model with chl-a (PS), lake TP and wind speed had the highest OOB score (Table 4). There are plausible mechanisms underpinning relationships between these three variables and lake chl-a, and all were selected for BN development. In the case of wind, windier weather can cause less stable lake stratification and lower chl-a concentrations (Huber et al., 2012; Yang et al., 2016). Air temperature exerted an important control on within-year changes in chl-a (Appendix A), but there was no evidence that years with higher summer air temperature were associated with higher mean chl-a concentration (Fig. 4).

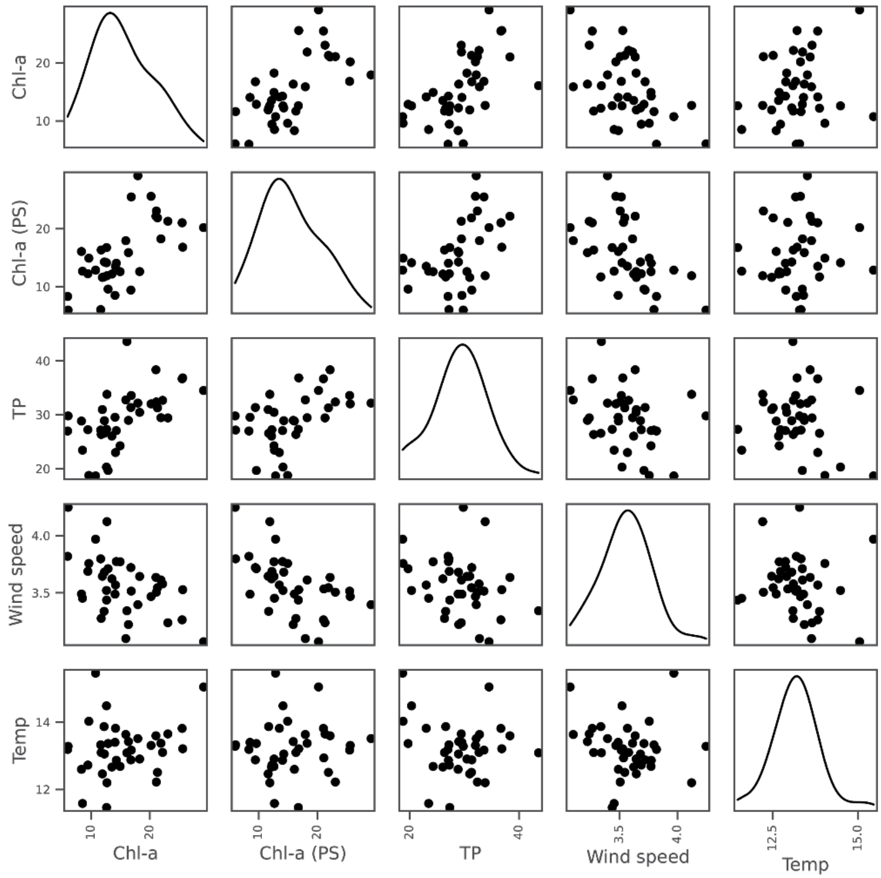


Figure 4: Relationships between seasonal mean chl-a (mg/l) and potential explanatory variables of interest, including chl-a from the previous summer (PS), seasonal mean TP (µg/l), wind speed (m/s) and air temperature (°C). Density plots estimated using kde are shown along the diagonal.

For cyanobacteria, the strongest correlation was with lake chl-a, although a number of other correlations were present (Table 3, Fig. 5). Feature importance analysis also highlighted chl-a as the most important variable (Table 4). Highest OOB values were obtained using just chl-a and lake colour, and these were therefore selected as the key explanatory variables for cyanobacteria. The relationship with lake colour is plausible, as an increase in organic matter can affect lake algal communities by reducing light availability and the availability of nutrients (Nagai et al., 2006), and Senar et al. (2021) found that above DOC concentrations of 8-12 mg/l, similar to those observed in lake Vanemfjorden (7-10 mg/l over the period 1996-2018), cyanobacteria became replaced by mixotrophic species as lake colour increased.

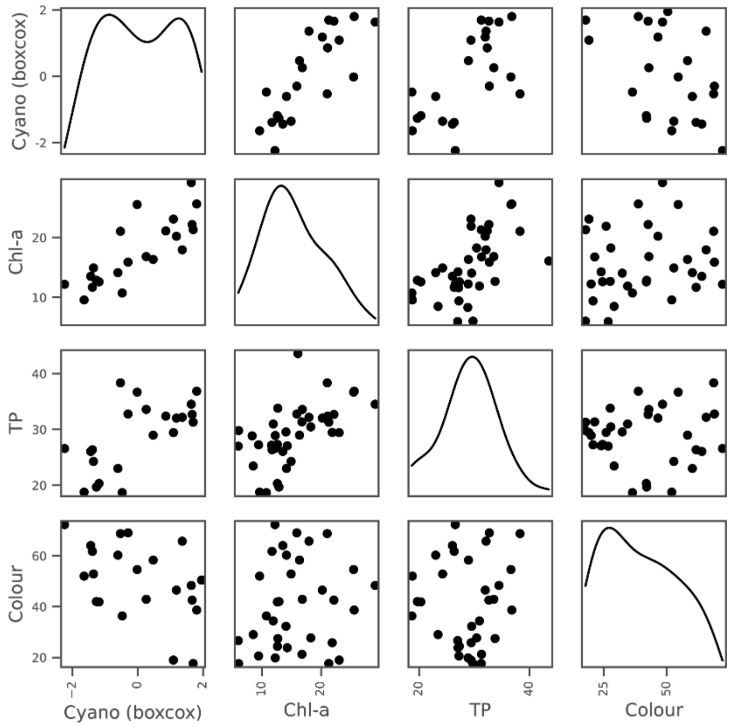


Figure 5: Relationships between Box Cox transformed maximum seasonal cyanobacteria biovolume and potential explanatory variables of interest, including seasonal means in lake chl-a, TP and colour. Density plots estimated using kde are shown along the diagonal.

Lake colour was very strongly correlated with the previous summer's colour (colour (PS)) and, probably because of this, the OOB score for lake colour was the highest of all the variables. Colour was also moderately correlated with factors relating to catchment delivery (Table 3, Fig. 6). The best regressor model had 3 features, including the previous summer's colour, calm wind days (wind < P20) and rain sum, although the latter two had low importance scores compared to the previous summer's colour (Table 4). As with TP, we suspect that the wind – colour relationship is not causative, as lake colour is relatively uniform throughout the water column in Vansjø so the impact of wind on lake stratification should be minimal. Wind was therefore dropped, and only the previous summer's colour and rain sum were selected.

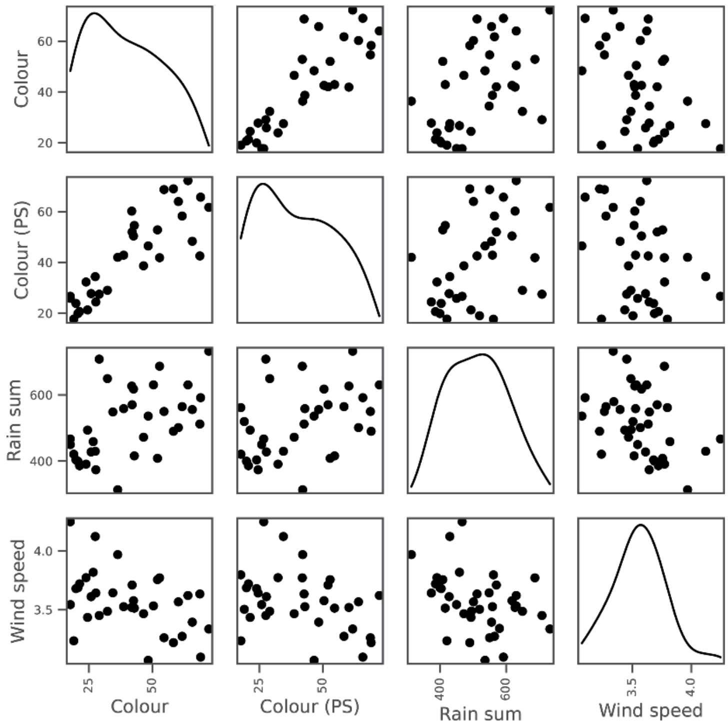


Figure 6: Relationships between seasonal mean lake colour and potential explanatory variables of interest, including colour the previous summer (PS), seasonal rain sum and mean wind speed. Density plots estimated using kde are shown along the diagonal.

In summary, the following features were selected for BN development for the four target variables:

- TP: lake TP concentration from the previous summer.
- Chl-a: chl-a from the previous summer, lake TP concentration, wind speed.
- Cyanobacteria: lake chl-a and colour.
- Colour: lake colour from the previous summer, precipitation.

3.2. Gaussian Bayesian network development

3.2.1. BN structure and GBN parameters

The key relationships highlighted (Section 3.1) were then used to develop the BN structure, which is shown, together with fitted coefficients for the GBN, in Fig. 7. For parentless nodes, coefficients define normal distributions with mean β_0 and variance σ^2. Child nodes are linear combinations of the parent nodes with intercept β_0, coefficients β_n and variance σ^2. Fitted coefficients for the Gaussian BN were all plausible, and matched the simple bivariate relationships between variables seen in the exploratory data analysis.

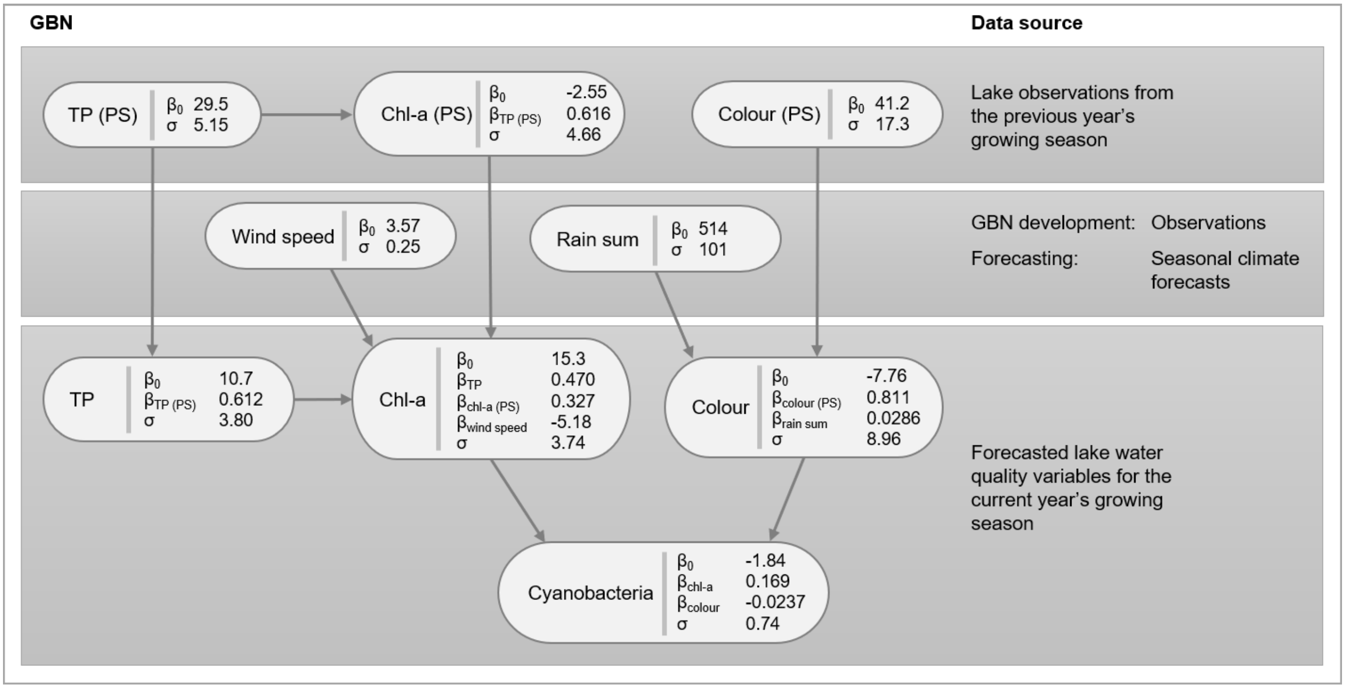
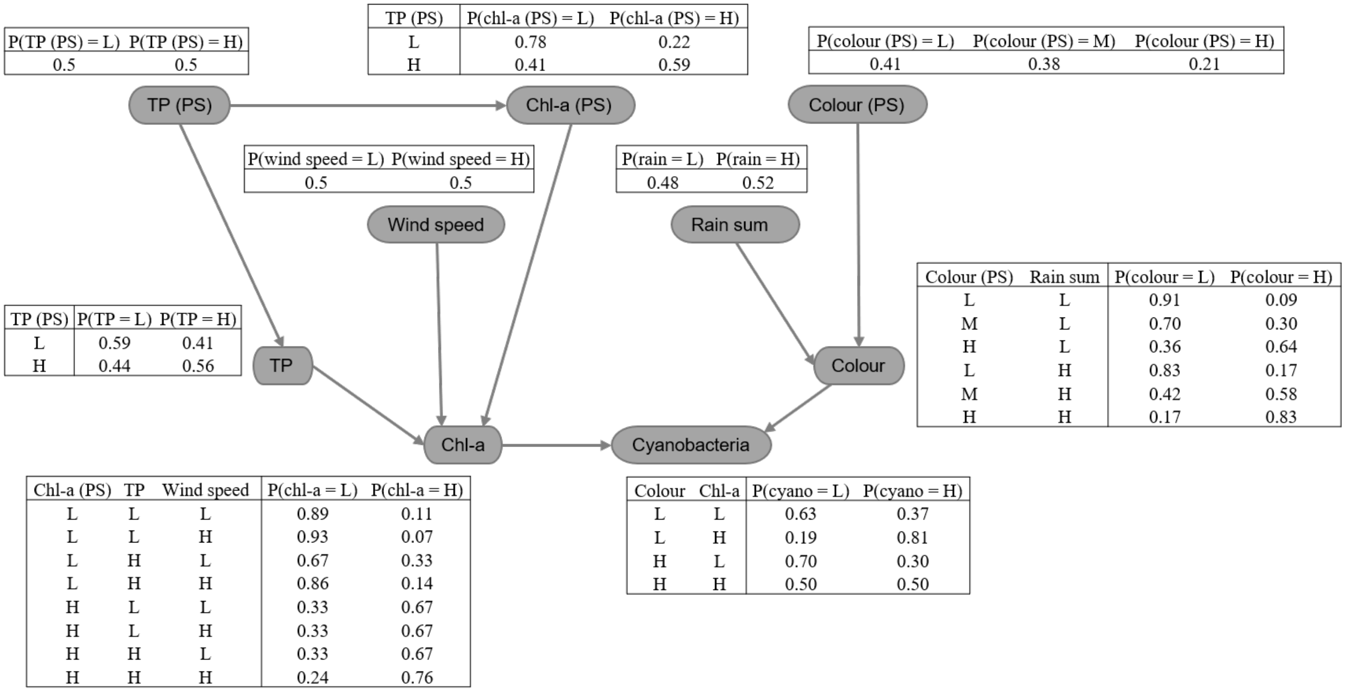


Figure 7: Gaussian Bayesian Network (GBN) structure and parameters defining the conditional probability densities at each node. Units for standard deviations (σ) and intercepts (β_0) are the same as the original data aside from cyanobacteria, where a box cox transformation was used (λ = 0.1). See Table 1 for a detailed description of the variables and Table B1 for 95% confidence intervals on the fitted coefficients.

3.2.2. Fitted discrete BN

The fitted CPTs for the discrete network (Fig. 8) did a slightly more mixed job of representing the relationships between variables. Despite using a relatively high iss value when fitting the network (i.e. giving the priors relatively high weight, see Section 2.62), several dubious relationships remain in the CPTs. For example, we expected a negative or no wind effect on chl-a, but in the last two rows of the chl-a CPT the opposite effect is seen, with an increase in the chance of having high chl-a at higher wind speeds.



P(TP (PS) = L)	P(TP (PS) = H)
0.5	0.5

TP (PS)	P(chl-a (PS) = L)	P(chl-a (PS) = H)
L	0.78	0.22
H	0.41	0.59

P(colour (PS) = L)	P(colour (PS) = M)	P(colour (PS) = H)
0.41	0.38	0.21

P(wind speed = L)	P(wind speed = H)
0.5	0.5

P(rain = L)	P(rain = H)
0.48	0.52

TP (PS)	P(TP = L)	P(TP = H)
L	0.59	0.41
H	0.44	0.56

Colour (PS)	Rain sum	P(colour = L)	P(colour = H)
L	L	0.91	0.09
M	L	0.70	0.30
H	L	0.36	0.64
L	H	0.83	0.17
M	H	0.42	0.58
H	H	0.17	0.83

Chl-a (PS)	TP	Wind speed	P(chl-a = L)	P(chl-a = H)
L	L	L	0.89	0.11
L	L	H	0.93	0.07
L	H	L	0.67	0.33
L	H	H	0.86	0.14
H	L	L	0.33	0.67
H	L	H	0.33	0.67
H	H	L	0.33	0.67
H	H	H	0.24	0.76

Colour	Chl-a	P(cyano = L)	P(cyano = H)
L	L	0.63	0.37
L	H	0.19	0.81
H	L	0.70	0.30
H	H	0.50	0.50

Figure 8: Fitted conditional probability tables for the discrete Bayesian network. Values were discretized into low (L) or high (H) classes (a medium (M) class was also included for colour (PS)) as described in Section 2.6.2.

3.3. GBN validation and assessment

We then explored the most appropriate GBN model structure and assessed its predictive performance using: (1) cross-validation using sub-sets of the GBN, including a comparison to the discrete BN, (2) goodness of fit of the whole network compared to observations, and (3) comparison to a simple benchmark model.

3.3.1. Cross-validation using sub-sets of the network

As mentioned in Section 2.7, cross-validation (CV) requires complete data for all variables and years. As cyanobacteria was only monitored since 1996, to avoid a large loss of data for TP, chl-a and colour, we split the GBN up into smaller sub-networks before performing cross-validation for each target node separately, as follows:

1. TP and chl-a: drop cyanobacteria, colour, previous summer's colour and rain nodes from the BN, and use the whole 1981-2018 period in cross-validation.
2. Colour: as colour was linked to the network through cyanobacteria, to be able to include the full period 1981-2018 we had to drop all nodes aside from colour and its parents.
3. Cyanobacteria: the whole network was used, but only with data from 1997.

CV results comparing the classification error of the GBN and the discrete BN are shown in Table 7. We might expect the discrete BN, which was fitted to discrete data, to do a better job of predicting the water quality class than the GBN. This was the case for all but TP, although it was only marginally better than the GBN for colour and cyanobacteria.

Predictive performance of the GBN with and without weather nodes is also shown in Table 5. Lake colour was the only variable for which model performance was a little better when meteorological variables were included, although the gains were marginal. For chl-a and cyanobacteria, performance was similar with or without weather nodes, and it was identical for TP. Overall, CV results suggest a marginal benefit to using precipitation when predicting lake colour, but that wind should be dropped from the GBN.

Table 5: Mean predictive performance of different Bayesian network (BN) structures, including the Gaussian Bayesian network (GBN) with and without weather nodes and a discrete BN, assessed through cross-validation. The BNs used to make predictions for each target variable were sub-sets of the BN shown in Fig. 7 for all but cyanobacteria, to make the most of all available data (see text). GBN cyanobacteria predictions were back-transformed to the original data scale before calculating statistics. Note: RMSE is root mean square error, NA is not applicable.

Variable	BN type	Weather nodes included?	R^2	RMSE	Classification error (%)
TP	GBN	✓	0.33	3.96	33
TP	GBN	✗	0.33	3.96	33
TP	Discrete	✓	NA	NA	41
Chl-a	GBN	✓	0.30	4.76	34
Chl-a	GBN	✗	0.29	4.76	32
Chl-a	Discrete	✓	NA	NA	8
Colour	GBN	✓	0.72	8.78	24
Colour	GBN	✗	0.68	9.35	24
Colour	Discrete	✓	NA	NA	15
Cyano	GBN	✓	0.14	1.91	31
Cyano	GBN	✗	0.22	1.76	31
Cyano	Discrete	✓	NA	NA	21

3.3.2. Goodness-of-fit of the whole network

Model performance of the whole network, assessed using the same data for fitting and assessment, is shown in Table 6. Performance was best for lake colour ($R^2 > 0.7$), which showed particularly high temporal autocorrelation. The same general lack of sensitivity to weather nodes was seen as in the CV results, and considering additional model performance (Table 6, Fig. 9).

Table 6: Performance of the GBN with and without weather nodes, fit using the whole historic period (no cross-validation) and the whole BN. Performance of the seasonal naïve forecast is also shown. MCC and classification error reflect classifier skill, whilst other statistics reflect how well mean predicted values matched observations. RMSE is root mean square error, MCC is Matthew's correlation coefficient.

Variable	Model	Weather variables included?	R^2	RMSE	Bias	MCC	Classification error (%)
TP	naïve	✗	0.40	4.39	0.49	0.18	41
	GBN	✓	0.42	3.67	-0.04	0.34	32
	GBN	✗	0.42	3.68	-0.07	0.34	32
Chl-a	naïve	✗	0.42	4.60	0.06	0.71	11
	GBN	✓	0.39	4.38	-0.08	0.23	27
	GBN	✗	0.37	4.44	-0.06	0.18	27
Colour	naïve	✗	0.72	9.21	0.85	0.55	21
	GBN	✓	0.75	8.39	-0.51	0.37	29
	GBN	✗	0.71	9.05	-0.75	0.44	26
Cyanobacteria	naïve	✗	0.32	1.76	0.18	0.57	22
	GBN	✓	0.36	1.53	-0.03	0.70	17
	GBN	✗	0.38	1.51	-0.01	0.70	17

3.3.3. GBN predictions compared to a benchmark model

Model performance was then compared to the performance of a seasonal naïve forecaster (Table 6, Fig. 9). For TP and cyanobacteria, the GBN performed slightly better than the naïve forecaster for all statistics. For lake colour, the GBN performed better at all but classification. For chl-a, the naïve forecaster performed slightly better, although this varied among performance statistics. It was particularly better at classification and, from inspection of Fig. 9, this is likely because the GBN predictions happen to often be just slightly under the 20 mg/l threshold used in classification.

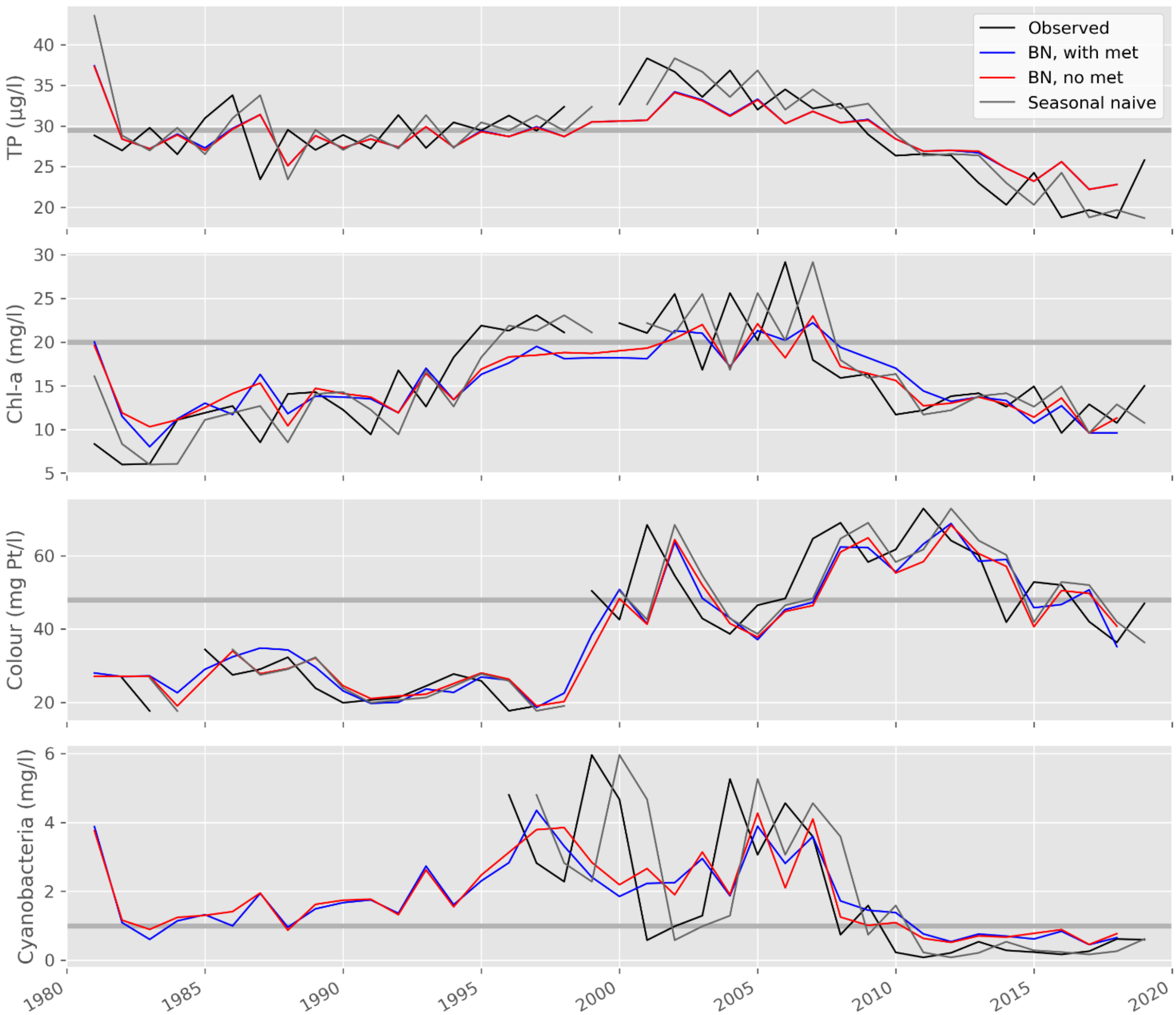


Figure 9: Lake water quality observations and predictions from a range of models, including the Gaussian Bayesian Network (BN) with and without weather variables and a seasonal naïve forecast. Horizontal grey lines show the thresholds used to classify predictions into two WFD-relevant classes (see Table 2).

3.4. Forecasting to support water management

An example of prototype seasonal forecasts, made using the GBN without weather nodes, is available at https://watexr.data.niva.no/ (last accessed 22/04/2022). The forecast includes the probability of being in one of two WFD-relevant status classes, the expected (mean) value, some historic skill information, and a text summary to aid in the interpretation of the forecast (e.g. "Chl-a is expected to be Moderate or better. Confidence level: Medium"). The forecast's layout was developed together with the region's water manager (Morsa) to ensure that it met their needs, and they have expressed optimism about the use of these kinds of forecasts to support water management, identifying actions which could be taken based on reliable-enough forecasts (Jackson-Blake et al., 2022). As well as providing an easy way of deriving probabilistic forecasts for both the expected value and the expected ecological status class, we found a real benefit of using BNs when co-developing models with stakeholders was the easy and transparent visualisation of the model used to derive the forecasts. We found that this increased stakeholder engagement with the model development process as well as their ability to correctly interpret the probabilistic predictions (Jackson-Blake et al., 2022).

4. Discussion

The main aims of this study were: (1) to develop a model for seasonal forecasting of lake water quality, and (2) to demonstrate the use of a continuous GBN for environmental modelling, instead of more traditional discrete BN approaches. We discuss each of these in turn below.

4.1. Seasonal forecasting of lake water quality

4.1.1. Drivers of interannual variability in lake water quality

In lake Vansjø, key water quality predictors were values observed the previous summer. Indeed, for lake TP concentration, this was the only predictor variable selected (Section 3.1). The strength of this annual autocorrelation, together with relatively low interannual variability in lake water quality (Fig. 9), are likely the reasons why the seasonal naïve forecast performed only slightly worse than the GBN and even slightly better for chl-a (Section 3.3.3). Aside from high temporal autocorrelation, we found positive relationships between lake TP concentration and chl-a and cyanobacteria, as widely documented elsewhere (Rousso et al., 2020). We also found a decrease in cyanobacteria as lake colour increased, again a previously documented effect (Section 3.1). No link was seen between lake colour and chl-a however, perhaps due to quality issues with the colour data before 1998 (Section 2.3), whilst cyanobacteria data were only available from 1996 and so missed the colour step-change. Although we found some evidence for relationships between weather variables and water quality, subsequent analysis suggested it was not worth including weather nodes in the GBN, as the improvements in predictive performance were marginal (for lake colour) or absent (Section 3.3), and it is highly unlikely that the marginal improvements would still be seen after replacing real observed historical meteorological data with seasonal climate model hindcasts.

The findings were relatively robust to the temporal aggregation window: statistical analyses using a shorter and more causally-plausible temporal aggregation resulted in very similar relationships being highlighted (Appendix A). The exception was that higher rainfall and discharge may result in lower cyanobacteria peaks, probably due to flushing, a relationship which was not accounted for in the GBN using 6-monthly aggregation and a potential area for improvement.

The lack of a temperature effect on algal biomass or cyanobacteria is interesting, as we might expect warmer summers to be accompanied by more intense blooms. However, results fit with a number of studies which found that warming effects were minor compared to nutrient effects (Lürling et al., 2018; Robarts & Zohary, 1987), and that water column stability was a key driver of cyanobacteria dynamics in dimictic lakes (Taranu et al., 2012) with wind playing a more dominant role than seasonal air temperature (Huber et al., 2012; Yang et al., 2016). We did however find a strong air temperature effect on within-year variation in chl-a and to a lesser extent cyanobacteria (Appendix A), likely because within-year variability is large compared to intra-annual variability and follows a systematic seasonal pattern. When looking in more detail at some of the BN studies in which relationships were identified between air temperature and algal variables (Couture et al., 2018; Moe et al., 2019; Rigosi et al., 2015; Shan et al., 2019; Williams & Cole, 2013), the observations used to fit the BN were not annually aggregated, so both with- and between-year variability were included. This may be appropriate if the aim is to look at algal dynamics within a year. However, it may not be appropriate for predicting inter-annual variation or longer term prognoses. Although temperature is likely to be important in many areas, it seems likely that a number of studies will have over-estimated its importance, by assuming that within-year relationships between temperature and algal dynamics can be used to infer future algal responses to increases in summer temperature under climate change.

4.1.2. Operational forecasting using seasonal climate data

One of the original aims of the study was to explore whether the latest seasonal climate forecasting products could be used to support water management, by enabling improved seasonal water quality forecasting. However, as we did not find a strong sensitivity to seasonal climate, this aim became redundant. In systems which are more sensitive to seasonal climate, a next step would be to assess GBN predictive performance using seasonal climate model hindcasts when making predictions (as in Mercado-Bettín et al., 2021). A comparison of model forecasting skill using seasonal climate data vs observed weather data would then allow for an assessment of the value of seasonal climate data. Seasonal climate forecasts are probabilistic and should only be used to give a broad indication of the likely direction of change, often in terms of tercile probabilities (e.g. “there is a 60% chance that next summer will be windier than normal”). A hybrid BN would therefore be a good option, with discrete nodes for the seasonal climate variables.

4.1.3. Data limitations and potential for improvement

As with all data-driven models, the quality of our model strongly relies on the availability and quality of the data, and in this regard we see potential for a number of improvements:

- Although the lake has a long history of monitoring, the training dataset is very small for a data driven model (≤ 39 data points). The lake showed low inter-annual variability, with gradual changes over time and few extreme events. Statistical power in a multivariate analysis is therefore limited, but will increase as more data become available.

- Peaks in cyanobacteria were defined by a single point, as in WFD classification, using relatively low frequency monitoring. An improvement would be for this value to be calculated more robustly, for example from the mean of a number of consecutive highest points.
- We only used data from a single point in the lake, whilst lake water quality can have high spatial variability. In Vanemfjorden, for example, there were bathing bans in place from 2000-2007, and yet the cyanobacteria data from the monitoring point is not particularly high during this period. Remote sensing products could help address this issue, and are increasingly being used in cyanobacteria bloom prediction (Bertani et al., 2017; Stumpf et al., 2012).

Overall, GBN predictions are almost entirely reliant on conditions observed during the previous summer. Despite the short residence time of the lake, if TP concentrations are buffered by lake sediment P release, seasonal algal peaks are not temperature limited, and water column stability is relatively insensitive to seasonal wind and temperature (e.g. because the water column is regularly mixed), then this rather simple model may be appropriate. All these things are plausible in this shallow lake with a long history of eutrophication. However, it is also likely that our model was limited by the underlying data, as mentioned above, and, for cyanobacteria, by the 6-month temporal aggregation window used. As an example of the limitation of the model, any events which happened during the previous winter are not currently taken into account when making forecasts. However, there is a general consensus that flooding in winter 2000 caused a large input of TP to the lake and was responsible for the cyanobacterial blooms that occurred in subsequent years (Haande et al., 2011). Our bottom up approach to selecting variables to include in the model meant that, as we did not find a relationship between winter discharge and lake TP concentration, it was not included. Whilst this bottom up approach ensures that the model is not affected by pre-conceived (but potentially incorrect) beliefs, it also means that rarely observed but perhaps important relationships are not included. In this case, incorporating expert knowledge to decide on additional nodes to include in the BN and on coefficients to define CPDs, could increase the robustness of the BN at predicting out-of-sample conditions, in particular the impacts of extreme events. An alternative, albeit more time-consuming approach, could be to include process-based model simulations to increase the size of the training data, assuming a robust model could be set up. The BN could then be used as a meta-model, as has been done previously at the site in the context of longer-term climate and land use change studies (Couture et al., 2018; Moe et al., 2019). However, process-based lake models typically only predict chl-a, so cyanobacteria forecasts would still rely on data-derived empirical relationships or expert knowledge.

4.2. Continuous GBNs for environmental prediction

With GBNs, it is straightforward to produce probabilistic predictions for water quality variables of interest. Predicting the probability of reaching a management target such as a specific a WFD status class is also straightforward and of direct management relevance (Section 3.4) and, although not demonstrated here, it is easy to update the training dataset using new data. These features make the approach well-suited to forecasting. In terms of performance, our GBN was modest in its prediction abilities. As discussed above, performance was likely limited by the nature of the lake and the data available for training, but we believe the approach itself was highly promising, and would likely result in a more powerful forecasting tool in lakes or rivers which showed higher inter-annual variability and sensitivity to seasonal discharge and climate, or if used for forecasting at shorter timescales (within-year, for example).

We found that perhaps one of the main benefits of using a GBN over a discrete BN to be the speed with which a sensible network can be developed. Our GBN parameters could be easily fit in a physically-plausible way only using observed data, despite the small dataset and the need to transform the cyanobacteria data. Developing a comparable discrete BN was a much more subjective and time-consuming process, both the discretization of the data, and also deciding on the weighting of the uniform prior to try to ensure sensible CPTs (Section 3.2.2).

However, the GBN approach has limitations which may be problematic in some settings. The normality assumption may not be appropriate, nor may assuming linear relationships between variables. Although there was no clear evidence for non-linear relationships (Section 3.1), they are common in ecological pressure-response relationships, including cyanobacteria blooms (Solheim et al., 2008). Overall, better performance might have been achieved with less stringent parametric requirements. Non-parametric or semi-parametric BN development has received a considerable amount of attention in recent years (Marcot & Penman, 2019), with a number of promising developments (e.g. Boukabour & Masmoudi, 2020; Hanea et al., 2015; Masmoudi & Masmoudi, 2019) and we expect that non-parametric continuous BN algorithms will increasingly become available in commonly-used BN software. However, the simplicity of the normal approximation used in GBNs means they may remain a good first choice. For people who use BN software that cannot handle continuous nodes, a good alternative could be to make use of commonly-available functionality which allows the user to specify a continuous probability distribution for a node, and then this is discretized within the software.

GBNs have much in common with Multiple Linear Regression (MLR), where linear relationships and Gaussian error distributions are usually assumed, and which are also able to produce probabilistic predictions of continuous variables. Indeed, the local distributions in a GBN are ordinary-least-squares regressions, i.e. univariate MLR involving only root nodes that are ancestors of the output. Both GBN and MLR approaches have advantages and disadvantages when it comes to environmental modelling and forecasting. MLR models have the advantage that input datasets do not need to be normally distributed and they are typically easy to implement with standard software. MLR has been successfully applied to algal bloom forecasting, for example in Lake Erie (Ho & Michalak, 2017). Benefits of the BN approach include, for example, ease of predicting multiple explanatory variables, as was the focus here, where the interest was in forecasting more than just algal bloom risk. Indeed, perhaps the main strength of using a GBN over MLR is that GBNs provide a powerful visual representation of potentially complex interdependencies between variables. By providing a convenient way of defining and visualising a multivariate model, where different outputs depend on different explanatory variables, it becomes easier to explicitly incorporate domain knowledge into the model building process (such as which variables affect which other variables), as well as facilitating collaborative model development and communication of results (Section 3.4). Based on our experience in this study, we believe the process of constructing a GBN forces modellers to think about key relationships, and to consider more carefully common MLR pitfalls such as multicollinearity and omitted variable bias.

5. Conclusions

We developed a continuous GBN to produce probabilistic forecasts for average growing season (May-October) lake water quality (TP, chl-a and colour) and maximum cyanobacteria biovolume. The aim was to provide early warning, in spring of a given year, of the likely conditions for the coming season. This is, to our knowledge, one of the first continuous GBNs for water quality prediction, and one of few continuous BNs in environmental modelling more generally. Overall, we found the GBN approach to be well-suited to seasonal water quality forecasting. It is straightforward to produce probabilistic predictions, including the probability of lying within a WFD-relevant status class. The process of developing the GBN was substantially less time-consuming and subjective than developing a discrete BN, and the GBN could be sensibly parameterised just using observed data, despite the small dataset. Despite the parametric constraints of GBNs, their simplicity, together with the relative accessibility of BN software which includes GBN handling, means they are a good first choice for BN development, which we think should be considered more widely when data are continuous.

Although the GBN approach itself proved to be promising, we had more mixed success with forecasting seasonal (or inter-annual) lake water quality at our study site. Although our exploratory data analysis suggested that wind and precipitation exerted a control on interannual variability in lake water quality, these relationships were weak, and overall our lake showed low sensitivity to seasonal climate. Instead, the dominant source of predictability was simply the lake water quality observed the previous year. Because of this strong inertia, the GBN did not perform much better than a naïve seasonal forecast. Potential improvements, which could make the model more powerful at predicting seasonal water quality, include incorporating expert knowledge on the likely impacts of rare events into the BN structure and conditional probabilities, improving the quality of the training data, and expanding the training set using synthetic process-based model results. We found a much stronger weather control on within-year variability in lake water quality, and we envisage a more management-relevant forecasting tool could be developed by adapting the approach to forecast water quality at sub-annual time scales, or by applying it to forecast seasonal water quality of water bodies (rivers or lakes) that show higher interannual variability and sensitivity to seasonal climate.

Appendix A: Exploratory statistical analyses using finer-scale temporal aggregation

A1. Method

Temporal aggregation over the whole growing season is coarse and may miss causative relationships. We therefore also carried out finer-scale aggregation, to check and expand on the results obtained from the 6-monthly analyses. This finer-scale aggregation included:

(1) *Algal peaks and pre-peak conditions for explanatory variables*: For each year, we selected peak (maximum) values for chl-a and cyanobacteria. We then calculated, for each of chl-a and cyanobacteria, means or sums of the potential explanatory variables over 14, 30, 60 and 90 days pre-peak. By ensuring that the potential explanatory variables only include data from before the observed algal peak, this aggregation method should have more power to identify causative relationships, whilst still focusing on inter-annual variation.

(2) *Monthly aggregation*. A repeat of the exploratory statistical analysis (Section 2.5) using monthly data, to explore the causes of both within- and between-year variability.

A2. Results

A2.1 Algal peaks and pre-peak conditions for the explanatory variables

For chl-a, strongest relationships were seen with lake TP concentration and wind-related variables (Table A1), as in the analysis using 6-monthly aggregation. For cyanobacteria, strongest correlations were with lake TP and chl-a concentrations, and there was also a relationship with lake colour, as in the 6-monthly analysis. In contrast to the whole-seasonal analysis, relationships between cyanobacteria and variables relating to wetness and flow were seen for some temporal aggregation windows, suggesting that the larger the rainfall and river discharge (and the shorter the lake water residence time) over the preceding 30-60 days, the lower the cyanobacterial biomass. Overall, this analysis using a shorter and more causally-plausible temporal aggregation resulted in very similar features being selected as in the whole-season aggregation, with the exception that hydrology and residence time may play more of a role in cyanobacteria bloom development.

Table A1: Pearson's R correlation coefficients between seasonal maxima of chl-a and cyanobacteria and potential explanatory variables aggregated (mean or sum) over *n* days before the algal peak occurred. For clarity, only |R| > 0.20 are shown for chl-a and |R| > 0.30 for cyanobacteria.

Variable	Temporal aggregation over *n* days pre-peak							
	n = 14		n = 30		n = 60		n = 90	
Chl-a	Wind speed	-0.35	Wind speed	-0.24	Wind > P80	-0.31	Wind > P80	-0.32
	Wind > P80	-0.32	Wind > P80	-0.22	Wind speed	-0.25	Wind speed	-0.23
					Wind > P60	-0.23		
	TP	0.21	TP	0.21	TP	0.34	TP	0.36
	Wind < P40	0.23	Wind < P20	0.23				
	Wind < P20	0.27						
Cyano	Colour	-0.33	Rain days	-0.41	Rain days	-0.45	Colour	-0.41
	Q	-0.31	Rain sum	-0.36	Rain sum	-0.39		
			Q	-0.33	Colour	-0.38		
			Colour	-0.33				
	Chl-a	0.48	Chl-a	0.54	Chl-a	0.48	TP	0.51
	TP	0.71	TP	0.63	TP	0.61	Chl-a	0.55

A2.2 Monthly aggregation

For all variables, strongest relationships were with values observed the previous months and there were strong correlations between values observed the previous summer. As well as this strong temporal auto-correlation, potentially important relationships included:

- TP: As in the 6-monthly analysis, the strongest relationship was with the previous summer's TP (R = 0.45), and there were weak relationships with wind. For example, the calmer the previous winter or 2-6 months, the higher the TP (R = 0.30 or less, depending on the lag), and the windier the previous winter or 6 months, the lower the TP (R ≤ -0.22, depending on the lag). Stronger relationships were seen between TP and wind over the previous ≥ 2 months, rather than the previous or current month. Wind should have an immediate and relatively short-lived effect on TP via water column mixing, so this suggests that the relationship is not causative. Relationships with all other variables were weak (R < |0.16|).
- Chl-a: strongest relationships were with the current month's air temperature (R = 0.50) and related variables (lagged air temperature, number of days in the current or previous months with sub-zero temperatures). Relationships with all other variables were weaker (R < |0.35|).
- Cyanobacteria: strongest relationships were with chl-a concentration (R = 0.71), lake colour (R = -0.43), lake TP concentration (R = 0.41), the previous summer's cyanobacteria and TP concentrations (R = 0.39, R = 0.37, respectively) and winter wind (R = 0.36 or lower, depending on the wind percentile).
- Colour: As in the 6-monthly analysis, strongest correlations were with the previous summer's colour (R = 0.72) and with rain variables. In particular, with the precipitation sum and the number of intense rain days over the previous five or six months (R in the range 0.56 – 0.60), and with discharge sum the previous 3 months (R = 0.54). There

was also a negative correlation with air temperature in the current or previous 1-3 months (R in the range -0.51 to -0.44). All other correlations had R < |0.41|.

Overall, many of the same variables which were important in explaining inter-annual differences were highlighted as being important. However, a key difference is the appearance of a strong relationship between air temperature and chl-a concentration, discussed further in Section 4.1.

Appendix B

Table B1: Fitted GBN coefficients with 95% confidence intervals

GBN node	Coefficient	Value	95% confidence interval (±)	
			Original data units	%
TP	β_0	10.73	7.53	70
	$\beta_{TP\ (PS)}$	0.612	0.252	41
TP (PS)	β_0	29.5	1.7	6
Chl-a	β_0	15.3	25.2	165
	β_{TP}	0.47	0.30	64
	$\beta_{chl\text{-}a\ (PS)}$	0.327	0.302	92
	$\beta_{wind\ speed}$	-5.18	6.02	-116
Chl-a (PS)	β_0	-2.55	9.15	-359
	$\beta_{TP\ (PS)}$	0.616	0.306	50
Wind speed	β_0	3.57	0.08	2
Cyano	β_0	-1.84	1.94	-105
	$\beta_{chl\text{-}a}$	0.169	0.069	41
	β_{colour}	-0.0237	0.0241	-102
Colour (PS)	β_0	41.2	6.0	15
Colour	β_0	-7.76	16.04	-207
	$\beta_{colour\ (PS)}$	0.811	0.221	27
	$\beta_{rain\ sum}$	0.0286	0.0342	119
Rain sum	β_0	514.2	33.3	6

Code and data availability

Data and scripts are available at https://github.com/NIVANorge/seasonal_forecasting_watexr, within the 'Norway_Morsa' folder.

Author contribution

LJB conceptualised and carried out the analysis, with input on limnological process understanding from SH, SJM and FC, on Bayesian network development from SJM, and with machine learning and Python/R integration support from JES. LJB prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Funding: This work was carried out as part of the WATExR project, part of ERA4CS, an ERA-NET initiated by JPI Climate. The work was funded by the Research Council of Norway (Project 274208), with co-funding by the European Union (Grant 690462). Many thanks to the whole WATExR project team for fruitful collaboration and discussions on the topic of seasonal forecasting and method development. Thanks also to José-Luis Guerrero for assistance in sourcing meteorological data.

References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., & Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling & Software*, *26*(12), 1376-1388. https://doi.org/https://doi.org/10.1016/j.envsoft.2011.06.004

Barton, D. N., Saloranta, T., Moe, S. J., Eggestad, H. O., & Kuikka, S. (2008). Bayesian belief networks as a meta-modelling tool in integrated river basin management — Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. *Ecological Economics*, *66*(1), 91-104. https://doi.org/https://doi.org/10.1016/j.ecolecon.2008.02.012

Bergström, A.-K., & Karlsson, J. (2019). Light and nutrient control phytoplankton biomass responses to global change in northern lakes. *Global Change Biology*, *25*(6), 2021-2029. https://doi.org/https://doi.org/10.1111/gcb.14623

Bertani, I., Steger, C. E., Obenour, D. R., Fahnenstiel, G. L., Bridgeman, T. B., Johengen, T. H., Sayers, M. J., Shuchman, R. A., & Scavia, D. (2017). Tracking cyanobacteria blooms: Do different monitoring approaches tell the same story? *Science of the Total Environment*, *575*, 294-308.

Boukabour, S., & Masmoudi, A. (2020). Semiparametric Bayesian networks for continuous data. *Communications in Statistics - Theory and Methods*, 1-23. https://doi.org/10.1080/03610926.2020.1738486

Bruno Soares, M., & Dessai, S. (2016). Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, *137*(1), 89-103. https://doi.org/10.1007/s10584-016-1671-8

Carpenter, S. R., Cole, J. J., Kitchell, J. F., & Pace, M. L. (1998). Impact of dissolved organic carbon, phosphorus, and grazing on phytoplankton biomass and production in experimental lakes. *Limnology and Oceanography*, *43*(1), 73-80. https://doi.org/https://doi.org/10.4319/lo.1998.43.1.0073

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1-13.

Couture, R.-M., Moe, S. J., Lin, Y., Kaste, Ø., Haande, S., & Solheim, A. L. (2018). Simulating water quality and ecological status of Lake Vansjø, Norway, under land-use and climate change by linking process-oriented models with a Bayesian network. *Science of the Total Environment*, *621*, 713-724.

Couture, R.-M., Tominaga, K., Starrfelt, J., Moe, S. J., Kaste, Ø., & Wright, R. F. (2014). Modelling phosphorus loading and algal blooms in a Nordic agricultural catchment-lake system under changing land-use and climate. *Environmental Science: Processes & Impacts*, *16*, 1588-1599.

D'Agostino, R., & Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b 2 and√ b. *Biometrika*, *60*(3), 613-622.

de Wit, H. A., Valinia, S., Weyhenmeyer, G. A., Futter, M. N., Kortelainen, P., Austnes, K., Hessen, D. O., Räike, A., Laudon, H., & Vuorenmaa, J. (2016). Current browning of surface waters will be further promoted by wetter climate. *Environmental Science & Technology Letters*, *3*(12), 430-435.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., & Stiassny, M. L. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), 163-182.

EC. (2003). *Overall approach to the classification of ecological status and ecological potential* (Water Framework Directive Common Implementation Strategy Working Group, Issue.

Geiger, D., & Heckerman, D. (1994). Learning gaussian networks. In *Uncertainty Proceedings 1994* (pp. 235-243). Elsevier.

Gozlan, R., Karimov, B., Zadereev, E., Kuznetsova, D., & Brucet, S. (2019). Status, trends, and future dynamics of freshwater ecosystems in Europe and Central Asia. *Inland Waters*, *9*(1), 78-94.

Gudimov, A., O'Connor, E., Dittrich, M., Jarjanazi, H., Palmer, M. E., Stainsby, E., Winter, J. G., Young, J. D., & Arhonditsis, G. B. (2012). Continuous Bayesian Network for Studying the Causal Links between Phosphorus Loading and Plankton Patterns in Lake Simcoe, Ontario, Canada. *Environmental Science & Technology*, *46*(13), 7283-7292. https://doi.org/10.1021/es300983r

Hanea, A., Morales Napoles, O., & Ababei, D. (2015). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety*, *144*, 265-284. https://doi.org/https://doi.org/10.1016/j.ress.2015.07.027

Hanlon, C. G. (1999). Relationships Between Total Phosphorus Concentrations, Sampling Frequency, and Wind Velocity in a Shallow, Polymictic Lake. *Lake and reservoir management*, *15*(1), 39-46. https://doi.org/10.1080/07438149909353950

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., Dortch, Q., Gobler, C. J., Heil, C. A., Humphries, E., Lewitus, A., Magnien, R., Marshall, H. G., Sellner, K., Stockwell, D. A., Stoecker, D. K., & Suddleson, M. (2008). Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae*, *8*(1), 3-13. https://doi.org/https://doi.org/10.1016/j.hal.2008.08.006

Ho, J. C., & Michalak, A. M. (2017). Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *Journal of Great Lakes Research*, *43*(3), 221-228. https://doi.org/https://doi.org/10.1016/j.jglr.2017.04.001

Huber, V., Wagner, C., Gerten, D., & Adrian, R. (2012). To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. *Oecologia*, *169*(1), 245-256.

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

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Haande, S., Solheim, A., Moe, J., & Brænden, R. (2011). Klassifisering av økologisk tilstand i elver og innsjøer i Vannområde Morsa iht. Vanndirektivet.

Ibelings, B. W., Fastner, J., Bormans, M., & Visser, P. M. (2016). Cyanobacterial blooms. Ecology, prevention, mitigation and control: Editorial to a CYANOCOST Special Issue. *Aquatic Ecology*, *50*(3), 327-331.

Jackson-Blake, L. A., Clayer, F., de Eyto, E., French, A. S., Frías, M. D., Mercado-Bettín, D., Moore, T., Puértolas, L., Poole, R., Rinke, K., Shikhani, M., van der Linden, L., & Marcé, R. (2022). Opportunities for seasonal forecasting to support water management outside the tropics. *Hydrol. Earth Syst. Sci.*, *26*(5), 1389-1406. https://doi.org/10.5194/hess-26-1389-2022

Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., & Lehikoinen, A. (2021). Bayesian Networks in Environmental Risk Assessment: A Review. *Integrated Environmental Assessment and Management*, *17*(1), 62-78. https://doi.org/https://doi.org/10.1002/ieam.4332

Kosten, S., Huszar, V. L. M., Bécares, E., Costa, L. S., van 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*(1), 118-126. https://doi.org/https://doi.org/10.1111/j.1365-2486.2011.02488.x

Kristensen, P., Whalley, C., Zal, F. N. N., & Christiansen, T. (2018). European waters assessment of status and pressures 2018. *EEA Report*(7/2018).

Lussana, C., Tveito, O. E., Dobler, A., & Tunheim, K. (2019). seNorge_2018, daily precipitation, and temperature datasets over Norway. *Earth Syst. Sci. Data*, *11*(4), 1531-1551. https://doi.org/10.5194/essd-11-1531-2019

Lürling, M., Mello, M. M. e., van Oosterhout, F., de Senerpont Domis, L., & Marinho, M. M. (2018). Response of Natural Cyanobacteria and Algae Assemblages to a Nutrient Pulse and Elevated Temperature [Original Research]. *Frontiers in Microbiology*, *9*(1851). https://doi.org/10.3389/fmicb.2018.01851

Marcot, B. G., & Penman, T. D. (2019). Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental Modelling & Software*, *111*, 386-393.

Masmoudi, K., & Masmoudi, A. (2019). A new class of continuous Bayesian networks. *International Journal of Approximate Reasoning*, *109*, 125-138. https://doi.org/https://doi.org/10.1016/j.ijar.2019.03.010

Matilainen, A., Vepsäläinen, M., & Sillanpää, M. (2010). Natural organic matter removal by coagulation during drinking water treatment: A review. *Advances in Colloid and Interface Science*, *159*(2), 189-197. https://doi.org/https://doi.org/10.1016/j.cis.2010.06.007

Mercado-Bettín, D., Clayer, F., Shikhani, M., Moore, T. N., Frías, M. D., Jackson-Blake, L., Sample, J., Iturbide, M., Herrera, S., & French, A. S. (2021). Forecasting water temperature in lakes and reservoirs using seasonal climate prediction. *Water Research*, 117286.

Merel, S., Walker, D., Chicana, R., Snyder, S., Baurès, E., & Thomas, O. (2013). State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. *Environment international*, *59*, 303-327.

Moe, S. J., Couture, R.-M., Haande, S., Lyche Solheim, A., & Jackson-Blake, L. (2019). Predicting Lake Quality for the Next Generation: Impacts of Catchment Management and Climatic Factors in a Probabilistic Model Framework. *Water*, *11*(9), 1767. https://www.mdpi.com/2073-4441/11/9/1767

Moe, S. J., Haande, S., & Couture, R.-M. (2016). Climate change, cyanobacteria blooms and ecological status of lakes: A Bayesian network approach. *Ecological Modelling*, *337*, 330-347.

Nagai, T., Imai, A., Matsushige, K., & Fukushima, T. (2006). Effect of iron complexation with dissolved organic matter on the growth of cyanobacteria in a eutrophic lake. *Aquatic microbial ecology*, *44*(3), 231-239.

Nojavan, F. A., Qian, S. S., & Stow, C. A. (2017). Comparative analysis of discretization methods in Bayesian networks. *Environmental Modelling & Software*, *87*, 64-71. https://doi.org/https://doi.org/10.1016/j.envsoft.2016.10.007

Paerl, H. W., & Huisman, J. (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental microbiology reports*, *1*(1), 27-37.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, *29*(3), 241-288.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Qian, S. S., & Miltner, R. J. (2015). A continuous variable Bayesian networks model for water quality modeling: A case study of setting nitrogen criterion for small rivers and streams in Ohio, USA. *Environmental Modelling & Software*, *69*, 14-22. https://doi.org/https://doi.org/10.1016/j.envsoft.2015.03.001

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., Kidd, K. A., MacCormack, T. J., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D., & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849-873. https://doi.org/https://doi.org/10.1111/brv.12480

Rigosi, A., Hanson, P., Hamilton, D. P., Hipsey, M., Rusak, J. A., Bois, J., Sparber, K., Chorus, I., Watkinson, A. J., & Qin, B. (2015). Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. *Ecological Applications*, *25*(1), 186-199.

Robarts, R. D., & Zohary, T. (1987). Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. *New Zealand Journal of Marine and Freshwater Research*, *21*(3), 391-399.

Rousso, B. Z., Bertone, E., Stewart, R., & Hamilton, D. P. (2020). A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*, *182*, 115959. https://doi.org/https://doi.org/10.1016/j.watres.2020.115959

Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *arXiv preprint arXiv:0908.3817*.

Scutari, M., & Ness, R. (2012). bnlearn: Bayesian network structure learning, parameter learning and inference. *R package version*, *3*.

Senar, O. E., Creed, I. F., & Trick, C. G. (2021). Lake browning may fuel phytoplankton biomass and trigger shifts in phytoplankton communities in temperate lakes. *Aquatic Sciences*, *83*(2), 21. https://doi.org/10.1007/s00027-021-00780-0

Shachter, R. D., & Kenley, C. R. (1989). Gaussian influence diagrams. *Management science*, *35*(5), 527-550.

Shan, K., Shang, M., Zhou, B., Li, L., Wang, X., Yang, H., & Song, L. (2019). Application of Bayesian network including Microcystis morphospecies for microcystin risk assessment in three cyanobacterial bloom-plagued lakes, China. *Harmful Algae*, *83*, 14-24.

Skarbøvik, E., Haande, S., Bechmann, M., & Skjelbred, B. (2021). Vannovervåking i Morsa 2020. Innsjøer, elver og bekker, november 2019-oktober 2020. *NIBIO Rapport*.

Solheim, A. L., Phillips, G., Drakare, S., Free, G., Järvinen, M., Skjelbred, B., Tierney, D., & Trodd, W. (2014). *Northern lake phytoplankton ecological assessment methods* (EUR 26503 EN). (JRC Technical Reports, Issue. E. Union.

Solheim, A. L., Rekolainen, S., Moe, S. J., Carvalho, L., Phillips, G., Ptacnik, R., Penning, W. E., Toth, L. G., O'Toole, C., Schartau, A.-K. L., & Hesthagen, T. (2008). Ecological threshold responses in European lakes and their applicability for the Water Framework Directive (WFD) implementation: synthesis of lakes results from the REBECCA project. *Aquatic Ecology*, *42*(2), 317-334. https://doi.org/10.1007/s10452-008-9188-5

Stumpf, R. P., Wynne, T. T., Baker, D. B., & Fahnenstiel, G. L. (2012). Interannual variability of cyanobacterial blooms in Lake Erie.

Søndergaard, M., Bjerring, R., & Jeppesen, E. (2013). Persistent internal phosphorus loading during summer in shallow eutrophic lakes. *Hydrobiologia*, *710*(1), 95-107. https://doi.org/10.1007/s10750-012-1091-3

Taranu, Z. E., Gregory-Eaves, I., Leavitt, P. R., Bunting, L., Buchaca, T., Catalan, J., Domaizon, I., Guilizzoni, P., Lami, A., & McGowan, S. (2015). Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecology letters*, *18*(4), 375-384.

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*(12), 3477-3490.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, *203*(3), 312-318. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2006.11.033

Vanndirektivet, D. (2018). *Klassifisering av miljøtilstand i vann*.

Welch, E. B., & Cooke, G. D. (2005). Internal phosphorus loading in shallow lakes: importance and control. *Lake and reservoir management*, *21*(2), 209-217.

Williams, B. J., & Cole, B. (2013). Mining monitored data for decision-making with a Bayesian network model. *Ecological Modelling*, *249*, 26-36. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2012.07.008

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839-2846. https://doi.org/https://doi.org/10.1016/j.patcog.2015.03.009

Yang, Y., Colom, W., Pierson, D., & Pettersson, K. (2016). Water column stability and summer phytoplankton dynamics in a temperate lake (Lake Erken, Sweden). *Inland Waters*, *6*(4), 499-508. https://doi.org/10.1080/IW-6.4.874